

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 20:58:35 ; Search time 5709 Seconds (without alignments)
16682.013 Million cell updates/sec

Title: US-09-806-767-1

Perfect score: 2328

Sequence: 1 atgggaacccaacattttttt.....taaccttgaaagactaactga 2328

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 segs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1:	gb_ba.*
2:	gb_hgt.*
3:	gb_in.*
4:	gb_om.*
5:	gb_ov.*
6:	gb_pat.*
7:	gb_ph.*
8:	gb_pl.*
9:	gb_pr.*
10:	gb_ro.*
11:	gb_sts.*
12:	gb_sy.*
13:	gb_un.*
14:	gb_vi.*
15:	em_ba.*
16:	em_fun.*
17:	em_hum.*
18:	em_in.*
19:	em_mu.*
20:	em_om.*
21:	em_or.*
22:	em_ov.*
23:	em_pat.*
24:	em_ph.*
25:	em_pl.*
26:	em_ro.*
27:	em_sts.*
28:	em_un.*
29:	em_vi.*
30:	em_htg_hum.*
31:	em_htg_inv.*
32:	em_htg_other.*
33:	em_htg_mus.*
34:	em_htg_pln.*
35:	em_htg_rod.*
36:	em_htg_mam.*
37:	em_htg_vrt.*
38:	em_sy.*
39:	em_htgo_hum.*
40:	em_htgo_mus.*
41:	em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB ID	Description
1	2328	100.0	78259	8	P20D22
2	691.2	29.7	150503	8	AC002411 Arabidops
3	482	20.7	2305	8	AC122149 Oryza sat
4	482	20.7	2618	8	AY142612 Arabidops
5	482	20.7	2686	8	AF360285 Arabidops
6	482	20.7	72698	8	AF065639 Arabidops
7	481.8	20.7	2265	6	AX506370 Sequence
8	481.8	20.7	2296	8	AY142613 Arabidops
9	481.8	20.7	2541	8	AY035090 Arabidops
10	480.4	20.6	2274	8	BT001082 Arabidops
11	480.4	20.6	2520	8	AY091773 Arabidops
12	456.6	19.6	2423	8	ATARA12 Arabidops
13	447.4	19.2	2549	8	AY080831 Arabidops
14	447.4	19.2	86212	8	AB010074 Arabidops
15	424.4	18.2	2525	8	AF360129 Arabidops
16	424.4	18.2	83906	8	AB022220 Arabidops
17	417	17.9	2638	8	AY084387 Arabidops
18	410.4	17.6	2645	8	AF436834 Arabidops
19	407.6	17.5	2650	8	LESBT2 Arabidops
20	407.6	17.5	5197	8	LES6379 Arabidops
21	398.2	17.1	2295	6	AX506491 Sequence
22	398.2	17.1	2326	8	AY096357 Arabidops
23	398.2	17.1	2665	8	AY074375 Arabidops
24	398.2	17.1	80346	8	ATM4E13 Arabidops
25	398.2	17.1	103150	8	ATF1111 Arabidops
26	398.2	17.1	195165	8	ATCHRIV82 Arabidops
27	397.2	17.1	2689	8	LESBT1 Arabidops
28	397.2	17.1	5089	8	LES6378 Arabidops
29	377.8	16.2	121160	8	AC005970 Arabidops
30	376.8	16.2	2465	8	BT002840 Arabidops
31	374.6	16.1	2152	8	PEU457185 Arabidops
32	365.4	15.7	154863	2	OSJN00071 Arabidops
33	359	15.4	199414	2	AC124955 Arabidops
34	354.6	15.2	128017	8	AC084282 Oryza sat
35	344.2	14.8	73859	8	AP000600 Arabidops
36	343.8	14.8	113049	8	AC134239 Arabidops
37	330.4	14.2	47295	8	F22M8 Arabidops
38	312.8	13.4	3173	8	LES17278 Arabidops
39	312	13.4	2680	8	LEAJ5172 Arabidops
40	311.4	13.4	3540	8	LES18932 Arabidops
41	311.2	13.4	2849	8	LEAJ5173 Arabidops
42	311	13.4	2407	8	LESUBTILI Arabidops
43	311	13.4	3076	8	LES17276 Arabidops
44	307.6	13.2	2847	8	LEAJ517 Arabidops
45	303.6	13.0	3975	8	LES18931 Arabidops

ALIGNMENTS

RESULT 1

F20D22

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

78259 bp DNA linear PIN 20-MAY-1998

Arabidopsis thaliana chromosome 1 BAC F20D22 sequence, complete sequence.

AC002411 GI:2570223

HTG.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 78259)

Vysotskaia.V.S., Osborne,B.I., Schwartz,J.R., Toriumi,M., Kwan,A.,

102(6)

TITLE	Yu,G., Oji,,O., Liu,S., Li,J., Hoang,L., Araujo,R. Au,M.,
JOURNAL	Brendel,V., Buehler,E., Conway,A.B., Conway,A.R., Dewar,K.,
REFERENCE	Feng,J., Kim,C., Kurtz,D., Li,Y., Palm,C.J., Shinn,P., Sun,H.,
AUTHORS	Davis,R.W., Ecker,J.R., Federspiel,N.A. and Theologis,A.
TITLE	Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence
JOURNAL	Unpublished (1998)
REFERENCE	2 (bases 1 to 78259)
AUTHORS	Theologis,A.
TITLE	Direct Submission
JOURNAL	Submitted (12-AUG-1997) Plant Gene Expression Center, 800 Buchanan
REFERENCE	Street, Albany, CA 94710, USA
AUTHORS	3 (bases 1 to 78259)
TITLE	Theologis,A.
JOURNAL	Direct Submission
REFERENCE	Submitted (28-OCT-1997) Plant Gene Expression Center, 800 Buchanan
AUTHORS	Street, Albany, CA 94710, USA
TITLE	4 (bases 1 to 78259)
JOURNAL	Theologis,A.
REFERENCE	Direct Submission
AUTHORS	Submitted (20-MAY-1998) Plant Gene Expression Center, 800 Buchanan
TITLE	St., Albany, CA 94710, USA
JOURNAL	On Oct 28, 1997 this sequence version replaced gi:2323245.
COMMENT	The sequence of BAC F20D22 from Arabidopsis thaliana chromosome 1. This sequence is of BAC F20D22 from Arabidopsis thaliana chromosome 1. The sequence of BAC F20D22 from Arabidopsis thaliana chromosome 1. 1. The sequence does not represent the sequence of the entire insert of this clone. It is shorter by 38223 bp because we sequence overlapping section only once. In order to facilitate the joining of overlapping clones in the future for creation of larger contigs, we provide a small overlap (200 bp) between overlapping submitted clones. The 5' end of this sequence overlaps by 200 bp the 3' end of the sequence of the BAC F19P19. Location/Qualifiers 1. .78259 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="1" /clone="F20D22" 3394..6024 /gene="F20D22.1" join(3394..4161,5206..6024) /note="F20D22.1" /gene="Strong similarity to beta-keto-CoA synthase gb U37088 from Simmondsia chinensis." /codon_start=1 /evidence=not_experimental /protein_id="AACI6740.1" /db_xref="GI:3142289" translation="MNENHITQSDHMNTIHVTKKLPFLSVLRKYVKLGHYHLINSSN AVYTLIPVLGLAATSGPSFLTDLLYNLHKFSLSTFAALLIFLTLTYPTTPR RTFFLDFACVPKPSSILICRETFEMDRSQVGITEFNLAFOOKILRSGLGOKTYP EALLRPDPNCMPARKEAETMFGADAVELEKTGVNPKDIGILVNCSLFNPFPIS AMIVPNKIRNVUSNLNGMGCSAGLISIDJAKQLQVQPNSALVSTENTILNYLS LGNDRSMLLNCSITFRMGAAVLNRSRSDRCRSKYLIHTVTRHGSDDDNAFCVORY EDNDNNKGIVSLSKNMATAAGEAKTNITLGPVLPMSEQLLFFATLVARKVFNYK KIKYVIPDEKLAFHFCTHAGGRAVLDEIEKNLDISEWHPESPMTLNRFGTSSSSSI WELPAYSEKGRJKEGDRTWQIAFGSGFKCNASAVRALRTIDPSKEKKKKTNPWDISI HEFPVPRTSFVTSSSESr complement(6494..11181) /gene="F20D22.2" complement(join(6494..6724,6808..6909,6996..7346, 7430..7654,7732..7888,8121..8302,8395..8824,8916..9655), 9777..9939,10056..10320,10685..11181)) /gene="F20D22.2" /note="Contains similarity to adenylate cyclase gb AF012921 from Magnaporthe grisea. EST gb Z24512 comes from this gene." /codon_start=1 /evidence=not_experimental /protein_id="AACI6742.1" /db_xref="GI:3142291" translation="MDSKIKKPAILEDADIDGGSESSTISSVLSLEDSDSVVDVSVC

NIEFSLDNDVDDSKGLYFRRNVFNLPKSI GGLGRRLKLFKFSNEIDLFPPELGNLV
 CHESLYQVKSIFGPGDGLSDWKLGKIKELBELTKVPKRSSALTLLISE SGLKCLRLSV
 NIEFYRLPEIICGLSELYLIDLSFKNPKIKGIPNEIGYLSLFTLKVAHRNLWELSPVL
 ALQNLLESLDNNRNLITLHPLDNLNMPRQLINLRNKLPSYCIWPTWICNFEGNY
 EEWGTVTSSSMVMVDVFETPYENNVTITPVHKGHRNPLANKSTGISLSRCSFSAKKS
 KRWKRRQYFQOQARQLRNLNRSKWEVPEGILSKMEVEETGKGWKVQNTDRGS
 VVNSDENDKCLFEASRVTTSBEESSLKADVSDNSQCVEZTQITSERDNGSEIKT
 SPFSQDAGPTVYDSESSRSGKPKNKRSKSEKYLDPNPKGCHKLSTIDNLSRKYS
 SNFGSTEDSLPGDFDAGDRPMTPEYKVELPLFSDREVLIDRAKDEVDLATAIS
 ARFALVAKKLNGLTPTDQVSYDNLQVSGFALFVSDFGSGDRTAIIERTKAVSG
 TNYQKPICTCLTNGNDLAALNKGVSTTBADAILDSDVCKSLRSIKSKRNSIYVPLG
 KQFICRRHALLKMYLCLDSMEPEVPCELVGYLDLDFMHANNIVPVKGSNVRMVD
 ACPHODIREDDQEFYCRYIPLNLRNRSIRKEKLEPGCSVSSLTSGKVERANSLL
 RCKLGSZAAVYKRWLTVEVSGLADITTFEYTCGLVEVRIKALGHKVDLVELYGHSEISS
 KWTTSSENGHVRQLTSSILMEHIRKGGSLKGI EKLSEAGKHGMDLARSYGLHSEISS
 LMLHNSJCTHROIKSNBNILIDNDSQNGSVIKLQCDRAVPLSHHGCSICARVG
 IPENICVGTPTMRSNPFVRAMHSONFVGLVEDTIVKSFCLTIFELLTLQNPYDLSBLQ
 IHSGLQGRKPKLPKLETLILSETEEBESNTKLSVEFDLITSDLDLTMFLIDVFHQCT
 EEPFSFDRINAGDLHEMILSRKRE
 complement (12002, .16201)
 /gene="F20022.3"
 complement (join(12002, .12091,12313, .12370,12485, .12576,
 12635, .12762,12905, .13016,13096, .13188,13287, .13377,
 13468, .13586,13763, .13876,13958, .14086,14252, .14426,
 14485, .14572,14649, .14894,15058, .15250,15291, .15577,
 15654, .15893,16012, .16201))
 /gene="F20022.3"
 /note="Contains similarity to gb|269902 from *C. elegans*."
 /codon_start=1
 /evidence=not experimental
 /protein_id="AAC16741.1"
 /db_xref="GI:3142290"
 /translation="WGGVSPFRKTGGDVSVAEYLFIATFVGEKSPFLASDFWNKLL
 LPLSRWSDRVQAOCELPFAGSNGVTRHLAKLILHSLCWLQLOAQDSSSLYKAAV
 NATYISVFLKYLLENGSKSDLSLHLSLDSDESPVGLSDQCNFVHVSFLVSGIG
 SNRYANADSLPFLICFSIRCEILLFIYILASPSNYVLHQELFMELMTSPGLQ
 LSGSHGTDANPITDAMTQLPHINFRHFLVNDVPEKSLVSLWRLLNLYISRH
 ILTPNKAISYSDSGSGLRVSASVALLPLPNLYIVNSNGSKNPLAESCSLHVLL
 RILNPKHSIMDSGQSLDERVSGSAVSESVKVFSDSDNTFKSALANARDVDRSDVE
 GNAPAGHVRI PASLFTDGLGILYILNLSMLSKLNNAIWOOFRLADEGAVLL
 YSLQSGDFEYVLVDPLDMLPLETLYANLSMRTSSNOIYMWLI VLLILSDSS
 FNSIHKRLSVSPWYKELHQSLGSLMWILIRTVQHNLKSLRDVYQTTCLATL
 ANAPAHHLISAYASQRLVLYMUSRKYNKLSLSDTUGKQVILKINLSDGVGSEDL
 AAELITDFLRLIDNALITLPAENPISUVAIMRQEVQFQFNHPRHELVEN
 IYVTCGIFTHREWSQVQLQFIINCRSWGEGGMKVNLYTYPSLKYLNLPFMFTQL
 HFSYEQSHPEFSEFRIYVWQLAFSRCSGFGFPDAILNFPVPHPEKEIEDERGESEG
 KAKVQELNEQRIVDP"
 17320, .19384
 /gene="F20022.4"
 complement (17320, .17477,17799, .17913,18007, .18057,18148, .18321,
 18430, .18609,18688, .18807,18887, .19020,19240, .19384)
 /gene="F20022.4"
 note="Contains similarity to tetraatrico peptide repeat
 protein gb|U46571 from home sapiens. EST gb|Z47802 and
 gb|Z48402 come from this gene."
 /codon_start=1
 /evidence=not experimental
 /protein_id="AAC16743.1"
 /db_xref="GI:3142292"
 /translation="MAEKAKATNGCBGEAKSLKEKGNPEFFKAGNELKAAALYTOAQL
 DFNSTALYSNRAAFSLVSLVKSALADAETTKINPQWFKVSIHONGIFRKGCVLEA
 MEKBDALAAEMALQNPQSTEVSRKIKRLGLOKQKQRAQLENLNFRKGVLEAKHLE
 SFKMSNENYGETECWKMPSTFVTMTKASWSEHETSXVDYRVPFLLDKEKTQTKY
 APANVIXAFQSPDHSNCTFYROYAEEFSKAACLVTPKSSISYVPMVKWGVGSRKW
 KLKPNGDGIFVFESESLRKLWFVSISSKEGQTLCRNRKARRALFCCSMILYIQFLIK
 TKKIKSCCECVLLSKSTFLLYLI
 complement (19889, .21254)
 /gene="F20022.5"
 complement (join(19889, .20488,20589, .21254))
 /gene="F20022.5"
 /note="Contains similarity to myosin IB heavy chain
 gb|X70400 from Gallus gallus."
 /codon_start=1

QY	304	GCAGTGAGACCTGACCAATGTTCTTCAGGTTCAACACCTTACTCTTTACAGTTCTTTGGGA	363
Db	86669	TCGGTTCGGGCGGACCCCGGGTTGAGCTGCACACGACGTACTCGTACAGGTTCTTTGGG	86728
QY	364	CTCAGCGGTTTGGAAACTCCGGTGTATGGTCTAAATCTCCGTTTGGTCAAGCCAAATT	423
Db	86729	CTCAGC---TTCTCTCCGACCGGCGCTGGGCGGTTCCGGGTACGGCGCGCACGATC	86785
QY	424	ATCGCGGTGCTTGATATCTGGAGTTTGGCTTGAAGTCTCTAGTCTTACGATACCGGAATG	483
Db	86786	ATCGCGGTGCTTCGACACCGGCGTGTGGCGGAGAAATCCCAAGCTTCGATGACCGGGATG	86845
QY	484	CTTTTCGATTCCACGGAATGGAAGGATTTGCCAAGAGAGAAAGTTTTCAGTTCTTCG	543
Db	86846	CCGCGGTGCGGCGCGTGGCAGGCGGTGTGCCAGGAGGTGAGCATTTCAACGCCACC	86905
QY	544	AGCTGTAAACCGGAAGCTAATCGGTGTAGATTCTTTCATCAGAGACACCGTGTTCGTAAT	603
Db	86906	AATTCGAATCGGAAGCTCATCGGCGCGGTTCTACTCCAAAGGCGCACCGCGCAACTAC	86965
QY	604	TCACCAGAGGAATCACCMAACATCCCTC---GTGAATACATTTCCGCAAGAGATTTCAACG	660
Db	86966	CCGACCAACCGTCCGACCGCGTGTGCTGATGAGTACGTGTGCGCGGGACGCGCAC	87025
QY	661	GGACACGGGACTCACACCGCTCAACAGTTGGTGGATCCTCTGTTTCGATGGCGAATGTT	720
Db	87026	GGGCACCGGAGCCACACGGCGTGCACGCGCCCGCGCGCGCGTGGCGCGCGCCAGCGTC	87085
QY	721	CTTGGCAATGAGCTGTGTGGCTCTGTGGATGCTCTCTCGAGCTCACATGTCAGTCTAT	780
Db	87086	CTGGCGGTGCGGGCGCGGACGCGCGCGGCGTGGCGCGCGCGCGCGCGCTAC	87145
QY	781	AAAGTCTCTGTTTCAATGGTTGTTTACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCG	840
Db	87146	AAAGTGTCTGTTTCAACGGGTGCTACAGCTCCGACATCTTCGCGGGGATGGACGACGCC	87205
QY	841	ATTCAAGATAAAGTCGATGTTCTTTTCGCTTTCCCTTGGCGGTTTCCCTATTCCTTTGAT	900
Db	87206	GTGCGCGAGCGGCTCGAGTGTCTGCTCTCCCTCTCCCTCGCGGCTTCCCATCCCGCTTTC	87265
QY	901	GATGACACATCGGCATTTGGAACATTCGAGCGCATGGAACCGGTATATCTGTAATCTGT	960
Db	87266	GAGGACAGCATCGGCATCGGAGCTTCCGTGCCACGACGATGGCGTCTCCGTCTGTGTC	87325
QY	961	GCAGCTGTGTAACACGGTCCCAATCGAAGCTCTGTGTCAAACACAGCTCTTGGGTCTCA	1020
Db	87326	GCGCGCGGGAACACGGGCGCATCGCGAGCTTCGTCGCCACGAGCGCCATGGGTGATC	87385
QY	1021	ACCATTTGCGCAGGACACGCTTGATCGAAGATTTCCCGCTGTGGTCAAGATTAGCCAACGGA	1080
Db	87386	ACCGTTCGGCGCGGACGCTGGACCGCGGATTCGCGGCTATGTCGCGCTCGGGAACGGC	87445
QY	1081	AAGCTTCTTATGGAGAGTCAATTGTATCCGGGAAAA---GGTATAAAGAAATGCCGGAGA	1137
Db	87446	CGGATCCTGTACGGCGAGTCCATGTTTCCCGGGAAGTTCGATTTGAAAAATGGCGGAAG	87505
QY	1138	GAGGTTGAGGTGATTTAGTTCACAGAGAGAGATAAGGAAGTGGTCTGTTTGGAGGG	1197
Db	87506	GAGCTCGAGCTGGTGTACCGCCAGTGGGACCGCGGAAGAGATGTACTGTCATCAAGGGG	87565
QY	1198	TCACTTCCAGAGAGAAATCCGAGCAAAATGGTGTATTGTGATCGCGGAGTCAATGGA	1257
Db	87566	GCCCTGTCTGCCGCCACTGTCTGCCGGAAGATGGTGTGTGTGACCGTGGCATCACCGGC	87625
QY	1258	AGATCGGAAAGGAGAGCGGTTAAAGAGCTGGAGAGTTGCAATGATCTTAGCCCAAT	1317
Db	87626	AGGGCTGACAAAGGCGAGGACGATTAACACAGCAGCGCGCGCCATGATTTCTCGCAT	87685
QY	1318	ACAGAGATCAACCAAGAGAAATCTATTGACGTTCTATTACCACTACATTTGATT	1377
Db	87686	TCCGAGATAAACCGAGAGAGACTCCGTCGATGTCCATGTCTCTCGCGCTCGACGCTCATC	87745
QY	1378	GGTTACACTGAGTCACTCTCTGAGGCTTATGTTAATGCCACGCGTGAACCAAGCGC	1437

Db	87746	GGCTACAGGAGGCGCTCGAGCTCAAGAACTACGTCAGCTCCACGCGCGCGCGGTGGCG	87805
QY	1438	CGGATAATTTTGGTGTATCGGTGATTTGGGAGGTCAAGACACCGAGGTGGCTCAGTTT	1497
Db	87806	AGGATAGTGTTCGGCGGACCGGATTTGGCGGGACCGCGCGCGGTGGCGCTGTTC	87865
QY	1498	TCAGCTCCAGGACCGAGTTTAGCCAAATCCTTTCGATACTTAAACCGGATATGATTGCTCCG	1557
Db	87866	TCCGCGCGCGCGGAGCTGACGAAACCGTCTGGTGTCTCAAGCCCGAGTGGTTCGCGCCCC	87925
QY	1558	GGAGTCAATATCATTTGCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTTATGAT	1617
Db	87926	GGCGTGAACATCATCGCGCTGGCCCGCAACCTGGGACCGTGGGCGCTGGAGGCGCAC	87985
QY	1618	TCAGAGAGAGTTAATTCATCTGTAATGTCAGGAACCTTCAATGTTCTTGTCCACATGTTAGC	1677
Db	87986	GCCCGCGGCTCCGACTTCAAGTGTCTCTCGGGACATCAATGGGTGCGCGCACGTGAGC	88045
QY	1678	GGAACTACTGCTCTTATCCGCTCTGCATACCGGAATCTGCTCTCAAGTGTCAATCAAAATCC	1737
Db	88046	GGCATCGCGCGCTCATCAGGTGCGCGCACCGCTCGTGGAGCGCGGATGGTGGCGTCC	88105
QY	1738	GCATTTGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAGAGGATAAAGATGGTAAC	1797
Db	88106	GGATCATGACGACCGCGGACGCTGACGACCGGCGGAGGCAAGCCGATCATGACGGGAAC	88165
QY	1798	AAACCA---GCCGCTGTGTTGCGATTTGGAGCAGGCGATGTGAATCCGCAAAAGCGGATA	1854
Db	88166	GGCGGCAAGGCGGACGCGTACGCCATGGCGCGCGGCACTGAACCCCGCGAGGCGCGTC	88225
QY	1855	AACCGGAGTTGGTTTACAACTTCAACAGTGGATTACATAAATCTCTGCACTCTT	1914
Db	88226	GACCGGCGCTCTGCTACGACATCGACCGGCGGCACTAGTCCACCACTCTGCACACTC	88285
QY	1915	GGATTCACAAGATCAGATATTTTAGCAATCACTCATAGAAACGTTGAGCTGCAATGGAATA	1974
Db	88286	GGGTACACCCACATGGAGATCTTCAAGATCACCGCTGGGCTCACTGACCGCGGTG	88345
QY	1975	TTGCGGAAAAACCCGGGTTTTAGTCTCAATTAACCGCTCGATAGCGGTGATTTTCAAACT	2034
Db	88346	CTGAGCGGAACCGCGCTTCAGCTCAACTACCCATCATATCGTGGCGCTTCAAGACG	88405
QY	2035	GGCAAGACTACGAGATGATCACAAGCGGTGTCACTAAGTGGGAGTCTTAACTCGATA	2094
Db	88406	AACACCACTCGCGGCTGTCTCCAGAGGACGCTCCCAACGTCGCGCACCGCACTCGACG	88465
QY	2095	TACTCAGTGAATGTCAAGCTCCAGAGGGATCAAGTTATTGTCAATCTTAAGAGACTT	2154
Db	88466	TACACCGCGAGGTCGCCCGCGCACCGGCTCAGGCTCAGGGTGTCCCGCGGACGCTG	88525
QY	2155	GTGTTCAAAACGTCGATCAGACGCTGAGCTATAGAT	2192
Db	88526	ACGTTCTCGAGTTTCGGCGAGAGAGAGCTTCCGGGT	88563

LOCUS	AY142612	2305 bp	mRNA linear	PLN 23-SEP-2002
DEFINITION	Arabidopsis thaliana putative subtilisin serine protease ARA12 (At5g67360) mRNA, complete cds.			
ACCESSION	AY142612			
VERSION	AY142612.1	GI:23296831		
KEYWORDS	FLI CDNA.			
SOURCE	Arabidopsis thaliana (thale cress)			
ORGANISM	Arabidopsis thaliana			
REFERENCE	1 (bases 1 to 2305)			
AUTHORS	Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R.,			

Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Arabidopsis Open Reading Frame (ORF) Clones

Unpublished

2 (bases 1 to 2305)

Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Direct Submission

Submitted (16-AUG-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEN (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEN) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEN) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

Location/Qualifiers

1. .2305

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="5"

/clone="U09277"

/note="This clone is in PUNI 51."

ecotype: Columbia"

1. .2305

/gene="At5g67360"

1. .2274

/gene="At5g67360"

/note="contains subtilase domains PF00082 and a PA domain PF02225"

/codon_start=1

/evidence=experimental

/product="putative subtilisin serine protease ARA12"

/protein_id="AA13181.1"

/db_xref="GI:23296832"

/translation="MSSEFLSSSTAFFILLCLGFCVSSSSDQGYIVHAKSQMPSSFDLHNSWYDSLRISDSAEILTYENAIHGFSTLTFQEEADSLMTQPGVIVLPPHRPEKVKGTVMCDGNARVQKGVVKAAGVGMILANTAAANGSELVADHLLPATTVIAEKAGDIIRHYVTDNPNTASISILGTVVGVSPVAAAFSSRGNSITPNILKPDLIA PGVNILAAWTGAQGTGLASDSRRVEENISGTSMSCPHYSGLAALLKSVHPWSPAAIRSAULTAYTKDKGLDLIATGKPSYSPVDLNPYFSAVNDGVGAYKTRIVTSVGFCLALNTYSPQIRSVSRNRYTCDPSKSYVADLNPYFSAVNDGVGAYKTRIVTSVCGAGTYSVKVTSSETTGKVISPEAVLNFKEANRKSYSVTFTVDSKSPSGNSPFGSIE2275. .2305

/gene="At5g67360"

BASE COUNT	547 a	567 c	580 g	611 t
ORIGIN				
Query Match	20.7%; Score 482; DB 8; Length 2305;			
Best Local Similarity	56.2%; Pred. No. 7.5e-126;			
Matches	974; Conservative 0; Mismatches 740; Indels 18; Gaps 3;			
Oy	210	ACTTCTCTACTCTATGGCTCTCGGATGAGGATTTGCTGCAGTTGACTGACTCAATCAGA	269	
Db	192	ATTGCTCTACACTTACGAGAACCGGATTCATGGAATCTCAACAGCGCTACTCAAGAAGA	251	
Oy	270	AGCGAGATACTAGATATTCACCTGAAAGTTGTCAGTGAGACCTGACCATTTCTTCA	329	
Db	252	AGCGACTCGCTCATGACTCAACTGGTGTTATCTCGTTTTACCGGAGCACCGTTACGA	311	
Oy	330	GGTTCAAAACCACTTACTCTTTACAAGTTCTTGGAGCTCGACGGTTTGGAAACTCCGGTGT	389	
Db	312	GCTACACACCACCTGCTACT	368	
Oy	390	ATGCTCTAAATCTCGGTTTGGTCAAGCACAATATTCGGGCTGCTTGCATCTGAGTTTG	449	
Db	369	GTTCCTCTGAAGCCGGCTCTTACAGCAGCTGCTCGGAGTTCTCGATACCGGAGTTTG	428	
Oy	450	GCCTGAAAGTCTTAGCTTTGACGATACCGGAATGCCCTTCGATTCCACGGAAATGGAAGG	509	
Db	429	CCCCGAGAGTAAAGCTTACTCCGACGAGGATTCGGTCTCTATTCCTTCTCTCTCTCTCT	488	
Oy	510	GATTTCGCAAGAGGAGAAAGTTTTCAGTTCTTTCGAGCTGTAAACGGGAAGCTAATTCGGTGC	569	
Db	489	CGGATGCGAGCCGCGAACCAATTTTACCGCTTCTCTCTGTAACCGTAATTAATTCGGAGC	548	
Oy	570	TAGATTCTTATCAGAGGACACCGTGTGCTGCTTAATTCACAGAGGAATCACCACATGCC	629	
Db	549	AAGATTCTTCGCTCGTGGTTAC-----GAATCAACGATGGGACCAATCGATGAATC	599	
Oy	630	TCGTGAATACATTTCCGCAAGAGATTCAACGGGACAGGGACTCACACCGCTCAACAGT	689	
Db	600	TAAAGATCAAGATCTCTAGAGACGACGCGTCAAGAACTCACCTCATCGACCGC	659	
Oy	690	TGGTGATCTCTGTTCGATGGGAAATGTTCTTTGGCAATGAGCTGGTGGCTCGTGG	749	
Db	660	CGCTGGATCCGTCGTTGAAGGAGCTAGCTTATTAGGCTACGCTTCAGGAACAGCTCGTGG	719	
Oy	750	GATGCTCTGGAGCTCATATGAGCTCTATAAGTCTGTGGTTCAATGTTGTTACAG	809	
Db	720	TATGCTCCACGCGCTGCTGCTGTTTACAAAGTCTGTGGCTCGGTGGTTGTTTTCAG	779	
Oy	810	CTCTGACATCTAGCAGCTATAGATGATAGATCAAGATAAGTCGATGTTCTTTCGCT	869	
Db	780	CTCAGATATTTTAGCTGCAATGATAAGCCATCGCGATACCGTCAATGCTCTCTCGAT	839	
Oy	870	TTCCCTTGGCGTTTCCCTATTCTTCTGATGATGACACAATCGCCATTGGAAATTCGG	929	
Db	840	GTCATCTGGCGGTGGAATGTCGGATATTATAGACCGGTGTTGGATCGGAGCATTCGC	899	
Oy	930	AGCCATGGAACCGGCTATATCTGTAATCTGTGAGCTGGTAAACAGCTGCAATCGAAAG	989	
Db	900	CGCCATGGAAGAGGGAATTTTAGTATCTTGTCTCAGCTGGTAAATGCTGCGAGTAGTTC	959	
Oy	990	CTCTGTGCAACACAGCTCTCTGGGTCTCAACCATTTGGCGGACGACGCTTGATCGAAG	1049	
Db	960	TAGTTTATCAACAGTAGTCTCATGGATCAACACTGTGGTCTGGTACTTTAGATCGTGA	1019	
Oy	1050	ATTTCCGCTGTGGTCAGATTAGCCACCGAAAGCTTCTATGTCAGAGTCAATCTATGCC	1109	
Db	1020	TTTTCCGGCGCTTGGCATTTCTCGGTAAACCGCAAGAAATTTACCGGAGTTTCTTTGTTAA	1079	
Oy	1110	GGCAAAAGGTATAAAGATTCGCGGAGAGAGGTTGAGGTGATTTACGTCAAGAGGAGA	1169	
Db	1080	AGGAGAGCTTTACCGGATAAATTTGTTGCCGTTTATTTACGCTGGAAATGCTAGTAATGC	1139	
Oy	1170	TAAAGGAAGTGAAGTTCTGTGTTGAGAGGGTCACTTCCAAAGAGAGAAATCCCGAGGCAAAAT	1229	

TITLE

JOURNAL

REFERENCE

AUTHORS

COMMENT

3' UTR

Db 1140 TACTAATGGTAATCTCTGTATGACCGGAACTTTGATCCGGAGAAAGTAAGGGGAAGAT 1199
Qy 1230 GGTGATTTGATCGCGAGTCAATGGAAGATCGGAGAAAGAGAGCGGTTHAAGAAC 1289
Db 1200 TGTGATGTGATAGAGAAATTAATGCTAGAGTTTCAAGAGGTGATGTGGTTAAGCAGC 1259
Qy 1290 TGGAGGAGTTCATGATCTTAGCCATACAGAGATCAACCAAGAGAGATCTATTGA 1349
Db 1260 AGTGGAGTTTGAATGATTTCTGGTAACACCTGGCGGGAATGGTGAAGAGCTTTTCGGGA 1319
Qy 1350 GGTTCATCTCTTACAGCTACATGATTTGGTTACACTGAGTCACTGCTTCTGAAGGCTTA 1409
Db 1320 TGCTCATTTGTTACCGGGACCCGTTGGTGAAGAGCGGTGATATATATCGGCATTA 1379
Qy 1410 TGTTAATGCCACCGTGAACCAAGGCGCGATTAATTTTGGTGGTACGGTGAATGGGAG 1469
Db 1380 CGTGACTACTGATCTTAATCCACCGCTTCGATTTCAATCTTAGGAACAGCTCGTCGGTGT 1439
Qy 1470 GTCAGGACCGGAGGTGGTCACTAGTTTTCAGCTCGAGGACCGAGTTAGCCATCCTTC 1529
Db 1440 TAAACCATCTCCGTTTGGCAGCGTTTAGCTACGTTGGACCGGAATTCGATTACACCGAA 1499
Qy 1530 GATCTAAACCGGATGATTTGCTCCGGAGTCAATATCATTCGCGCTTGGCTCAAAA 1589
Db 1500 TATTCCTTAAACCGGATGATTCGCTCTCGAGTAAACATCTCGCCGCTGGACCGGTGC 1559
Qy 1590 TCTAGGACCAACCGGACTTCTTATGATTTCAAGAGAGTTAACTTCACTGTAATGTCAGG 1649
Db 1560 TGTGGACCAACCGGACTCGTCTCGGATCTCGCGGCTGGAGTTCAATATCATCTCTGG 1619
Qy 1650 AACTTCATGTTTGTCCATGTTAGGGAATCACTGCTTCTATCCGCTCTGATACCC 1709
Db 1620 GACCTCGATGTTTGGCTCAGTTAGTGTGTTAGCGCGCTTCTCAAGTCGTTGATCC 1679
Qy 1710 GAATGCTCTCAGCTCAATCAATCGGATTCATGATGATGATGATGATGATGATGATG 1769
Db 1680 TGAATGGAGCCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 1739
Qy 1770 TCAAGGGAACCGATA-----AAGGATGGTAACAAACCGCGGTGTTGCGATTGG 1823
Db 1740 AGACGGTAAACGGTTACTGACATCGCGACAGGAGAGCTTCGACGCGTTCGATCAGG 1799
Qy 1824 AGCAGGCGATGTAATCGGAAAGGCGATAAACCAGGATGGTTTACAACTTCAACC 1883
Db 1800 TGCAGGACACGTTGACCAACCACTGCCACTATCCAGGACTATCCTACGATCTAACGAC 1859
Qy 1884 AGTGATTAACATTAATCTACCTCTGCACTCTTGGATTGATGATGATGATGATGATG 1935
Db 1860 GGAGGATTAATCTAGCTTCTCTGCGATGATGATGATGATGATGATGATGATGATGAT 1911

RESULT 4
AF360285 2618 bp mRNA linear PLN 18-SEP-2002
LOCUS Arabidopsis thaliana putative subtilisin serine protease ARA12
DEFINITION (At5g67360) mRNA, complete cds.
ACCESSION AF360285
VERSION AF360285.1 GI:3430745
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 2618)
Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Ban, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Full Length cDNA clones

TITLE

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 2618)
Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Ban, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission

TITLE
JOURNAL
COMMENT

Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-length cDNA': Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X.,
Sakano, H., Pham, P.K., Ban, J., Chung, M.K., Goldsmith, A.D.,
Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Chen, H.,
Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J.,
Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to Genbank.

FEATURES
source

Location/Qualifiers
1..2618
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFLO5-16-L13 (R09277)"
/note="this clone is in a modified pbluescript vector
(FLC-1) as a BamHI/XhoI insert."
ecotype: Columbia

gene

1..2618
/gene="At5g67360"
1..114
/gene="At5g67360"
115..2388
/gene="At5g67360"
/note="contains subtilase domains PF00082 and a PA domain
PF02225"

CDS

/codon_start=1
/evidence=experimental
/product="putative subtilisin serine protease ARA12"
/protein_id="AAK2595.1"
/db_xref="GI:3430746"
/translation="MSSSFLSSSTAFFLLCLGCFHVVSSSSDQGTIVVHAKSQMPSS
FDLHNSWYDSSLSISDAELLYTENAIGHSTLQEDASLMTQPGVSLVPEHR
VELHTRTPFLGLDEHADIPEAGSYDVVGVLDTCVWPEKSYSGDEGFGPIPS
WKGCEAGTNTASLCKNLICARFARGVESTWGPIDESKESRSPRDDGHTHTSS
TAAGSVGASLLGYAGTAGMADPARVAVYKVCWLGCCSSDLAIDKADNVN
VLSMLGGMSDYRDGVAIGAAAMERGIIVSCAGNAGSSSSLSNVAWITTVGA
GLDTRDFALAILNGKNTGVSPKGEALPDLPLFFIYAGNAGNATNGLCMTGLI
PKVKGTVMCDRGINARVQKGVVKAAGGVGMILANTAAAGELVADAHLLPATTVG
EXAGDIIRHYVTDDNPNTASILGTVVGVKPSVWAAFSRGPNSITPNTILKPDLIA
PGVNIILAWTGAAGTGLASDSRRVFEINIISGTSMSCPHVSLAALLKSVHPEKSPA
IKSALMTAYKTKDGLDIAATKPSFPDGHGAVSPPTATNPLGIYDLTIEDYL
GFLCALNTSPQIRSVRNKTCDEKSYVADLNYPFAVNVGVDGVAIKYKTRVTSV
GGAGYSVKVTSETTGTGKVISPEVPLNFKEANEKKSITVTITVDSKPSGNSFGSIE
WSDGKHVGVSPVAISWT"
2389..2618
/gene="At5g67360"
misc_difference 2602
/gene="At5g67360"

/note="not present in genomic sequence"

BASE COUNT
ORIGIN

642 a 632 c 628 g 716 t

Query Match 20.7% Score 482; DB 8; Length 2618;

Best Local Similarity 56.2%; Pred. No. 7.5e-126;

Matches 974; Conservative 0; Mismatches 740; Indels 18; Gaps 3;

Qy 210 ACTTCTCTACTCCTATGGCTCTCGGATTGAAGATTGTTGCTCAGTTGACTGAATCAGA 269
 Db 306 ATTGCTCTACACTTACGAGAACGGATTTCATGATTCTCAACACGGCTAACTCAAGAAGA 365
 Qy 270 AGCCGAGATCTAGATATTCACCTGAAGTTGTTGAGTCAGACCTGACCATGTTCTTCA 329
 Db 366 AGCCGACTCGCTCATGACTCAACCTGGTGTATCTCCGTTTACCGGAGCACCGTTACGA 425
 Qy 330 GGTTCACACACTTACTCTTACAGTTCTTGGGACTCGACGGTTTTGGAACTCCCGGTGT 389
 Db 426 GCTACACACACTCGTACT 482
 Qy 390 ATGGTCTAAATCTCGGTTTGGTCAAGGCACAAATTATCGCGCTGCTTGATCTCGAGTTG 449
 Db 483 GTTCCCTGAGCGGCTCTTACACGACGCTCGTCTCGGAGTTCTCGATACCGGAGTTTG 542
 Qy 450 GCCTGAAGTCTAGCTTTGAGGATACCGGAATGCCCTTCGATTCCACGGGAATCGAAAG 509
 Db 543 GCCCGAGATAAAGCTACTCCGACGAAGGATTCCGTCCTATTCTCTCTCTCTCTCTCT 602
 Qy 510 GATTGGCAAGAGAGAAATTCAGTTCTTTCGAGCTGTAAACCGGAAGCTATCCGTGC 569
 Db 603 CGGATCGGAGCGGAGAACAAATTTACCGGCTTCTCTCTGTATACGTTAACTATTCGAGC 662
 Qy 570 TAGATTCTTCTACAGACACACCGTCTCGTAAATTCACGAGGAATTCACCAACATGCC 629
 Db 663 AAGATTCTTCGCTCGTGTAC-----GATCAACGATGGACCAATCATGATGATC 713
 Qy 630 TCGTGAATACATTTCCGCAAGAGATCAACGGGACACGGGACTCACACCGCTCTACAGT 689
 Db 714 TAAAGATCAAGATCTCTAGAGACGACGACGGTCAACGGAATCACACCTCATCGACCGC 773
 Qy 690 TGGTGGATCCTCTGTTTCGATGGGAAATGTTCTTGGCAATGGAGCTGGTGGCTCGTGG 749
 Db 774 CGCTGGATCCGCTGTTGAAGAGTAGCTTATTAGGTACGCTTCAGGAACACGCTCGTGG 833
 Qy 750 GATGCTCCTGAGCTCATTGTCAGTCTATAAAGTCTGTGTTCAATGTTGTTACAG 809
 Db 834 TATGGCTCCACGCGCTGCTGCTGCTGTTTACAAAGTCTGTGGCTCGGTGTTTCAG 893
 Qy 810 CTCTGACATTTAGCAGCTATAGATGTAGCGATTCAAGATAAAGTCGATGTTCTTCGCT 869
 Db 894 CTCAGATATTTAGCTGCAATCGATAAAGCCATCGCCGATAACGTCATATGTCCTGTCAT 953
 Qy 870 TTCCTTCGGGTTTCCCTATTCCTTTGTATGATGACACAAATCCCAATTCGGAATTCGG 929
 Db 954 GTCACTCGCGGTGGAATGTCGATTTATAGAGACGGTGTTCGATCGGAGCATTCGC 1013
 Qy 930 AGCCATGGAACCGGTATCTGTAACTGTGCGAGCTGTAAACACGCTCCCAATCGAAAG 989
 Db 1014 CGCCATGGAAGAGAGGATTTTAGTATCTTGTCTAGCTGTGTAATGCTGTCGAGTAGTTC 1073
 Qy 990 CTCTGTTGCAACACAGCTCCTTGGTCTCAACCATTTGGCGCAGCGACGTTGATCGAAG 1049
 Db 1074 TAGTTTATCAACGTAGCTCATGATCACTGTTGCTGCTGTTAGATCTGTA 1133
 Qy 1050 ATTTCCCGTGTGGTCAAGTATAGCCAGGAAGCTTCTATGGAGAGTCAATGTATPC 1109
 Db 1134 TTTTCGGCGCTTGGATTCTCGGTACCGGAAGAAATTCACCGGAGTTTCTTGTGTTAA 1193
 Qy 1110 GGGAAAGGTATAAAGATCCCGGAGAGAGTTGAGTGATTACCTCACAGGAGAGA 1169
 Db 1194 AGGAGAAGCTTTACCGGATAAATTTGTCGGTTATTTACGCTGGGAATGCTAGTAATGC 1253
 Qy 1170 TAAAGGAGTGTGTTCTGTTTGGAGGGTCTACTTCCAGAGAGAAGAAATCCGAGGCAAAAT 1229

Db 1254 TACTAATGGTAAATCTCTGTATGACCGGAACCTTTGATCCCGAGAAAGTAAAGGGAGAT 1313
 Qy 1230 GGTGATTGTGATCGCGGAGTCAATGGAAGATCGGAGAAAGGAGACGGTTAAAGAGC 1289
 Db 1314 TGTGATGTGTGATAGAGAAATTAATGCTAGAGTTCAAGAAAGTGTGTGTTAAAGAGC 1373
 Qy 1290 TGGAGGAGTTTCAATGATCTTAGCCAATACAGAGATCAACCAAGAAAGAGATTCTATTGA 1349
 Db 1374 AGGTGGAGTTGGATGATCTGGCTAACACTCGCGCGAATGGTGAAGAGCTTGTTCGGA 1433
 Qy 1350 CGTTCACTCTTTACCACTACATTTGATTGTTTACCTGAGTCAGTCTCTCTGAAAGGCTTA 1409
 Db 1434 TGTCTAATTTGTTTACCGGACACCGTTGTGTAAGAAAGCCGGTGATATAATCCGCAATTA 1493
 Qy 1410 TGTTAATGCCACGGTGAACCAAGAGCGCGATTAATTTTGGTGGTACGGTGTGAGAG 1469
 Db 1494 CGTCACTACTGATCTTAATCCACCGCTTCGATTCAATCTTAGAACAAGTCGTCGGTGT 1553
 Qy 1470 GTCAGGACACCGGAGTGGCTCAGTTTTCAGCTCGAGGACCGAGTTTAGCCAATCCCTTC 1529
 Db 1554 TAAACCATCTCCGTTTTCGAGCGTTTAGCTACGTGGACCGAATTCGATTACACCGAA 1613
 Qy 1530 GATACATAAACCGGATGATTGTCGCGGAGTCAATATCAATTCGCGGTTGGCCCTCAAAA 1589
 Db 1614 TATTCTTAAACCGGATCTGATCGCTCTCGAGTAAACATCTCTCGCGCTGGACCGTGC 1673
 Qy 1590 TCTAGACCAACCGGACTTCTTATGATTCAAGAGAGTTAACTTCACTGTAATGTCAGG 1649
 Db 1674 TGTGGAACCAACCGGACTCGCTTCGATTCTCGCGCGTGGAGTTCAATATCATCTCTGG 1733
 Qy 1650 AACTTCAATGTTGTTCCACATGTTAGCGGAATCACTGCTCTTATCCGGTCTGCATACC 1709
 Db 1734 GAGTCGATGTTGTCCTCCTCAGTTAGTGGTTTAGCGCGCTTCTCAAGTCGGTGCATCC 1793
 Qy 1710 GAACCTGGTCTCCAGTGCATCAATCCGATTCGATGACAAACAGCGGATTTGTACGATCG 1769
 Db 1794 TGAATGGAGCCGCGCGGATTAGATCGCGCTTATGACCAACCGCTTACAAACCTACAA 1853
 Qy 1770 TCAAGGAAAGCGATA-----AAGGATGTTAAACAAACACCGCGGTGTTGGATGG 1823
 Db 1854 AGACGTTAAACCGTTTACTCGACATCGACAGGAAAGCTTCGACCGCTTCGATCAGG 1913
 Qy 1824 ACAGCGGATGTAATTCGCAAGAGCGGATTAACCCGGATTTGTTTACACATTCACAC 1883
 Db 1914 TGCAGGACACGTTGTCACCAACAACTGCCAATATCCAGGACTCATCTACGATCTAACGAC 1973
 Qy 1884 AGTGGATTACATAACTTACCTCTGCACTCTTGGATTCAAGATCAGATATT 1935
 Db 1974 GGAGATTACTTAGGCTTCCTCTCGCATTTGAATTACACATCGCCGCAAAAT 2025

RESULT 5

AF065639

LOCUS

DEFINITION

Arabidopsis thaliana cucumisin-like serine protease (ARA12) gene,

complete cds.

AF065639

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AF065639 2686 bp DNA linear PIN 14-AUG-2000
 Arabidopsis thaliana cucumisin-like serine protease (ARA12) gene,
 complete cds.

AF065639.1 GI:3176873

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 2686)

Ribeiro, A., Akkermans, A. D., van Kammen, A., Bisseling, T. and

Pawlowski, K.

A nodule-specific gene encoding a subtilisin-like protease is
 expressed in early stages of actinorhizal nodule development

Plant Cell 7 (6), 785-794 (1995)

95375542

7647567

REFERENCE 2 (bases 1 to 2686)
 AUTHORS Yamagata,H., Uesugi,M., Saka,K., Iwasaki,T. and Aizono,Y.
 TITLE Molecular cloning and characterization of a cDNA and a gene for subtilisin-like serine proteases from rice (*Oryza sativa* L.) and Arabidopsis thaliana
 JOURNAL Biosci. Biotechnol. Biochem. 64 (9), 1947-1957 (2000)
 MEDLINE 20505592
 PUBMED 11053401
 REFERENCE 3 (bases 1 to 2686)
 AUTHORS Yamagata,H.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAY-1998) Laboratory of Biochemistry, Faculty of Agriculture, Kobe University, Nada, Kobe 657-8501, Japan
 FEATURES
 Location/Qualifiers
 1..2686
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 <290..>2563
 /gene="ARA12"
 <290..>2563
 /gene="ARA12"
 /product="cucumis-in-like serine protease"
 290..2563
 /gene="ARA12"
 /note="similar to subtilisin-like protein encoded by GenBank Accession Number X85974; intronless gene"
 /codon_start=1
 /product="cucumis-in-like serine protease"
 /protein_id="AAC18851.1"
 /db_xref="GI:13176874"
 /translations="MSSSFLSSTAFFLLLLGFCVSSSSDQGTIVVHMAKSPMPSS
 PDLHNMVDSLSISDSABELLYENAIHFSTRLTQBEADSLMTPQGVISVLPFHR
 YELHTPTPLFLGDSHTADLFPKAGSYDVVGVLDGTGWPESKSYSGEGPPTPS
 WKGCERAGFTSLNKRKLIIGARFPARGVSTMGIDLSKSRSPRDDGHTHTSS
 TAAGSVVSGASLLIGASGTARGAPARVAIVKVGCGFCSDILLADIKAIADNVN
 VLSNLSGDSMDYRDGVATGAFAEMERGLVSCSAGNAPSSSSLSNVAPWMTTGA
 GTLDRDPALALINGNFTGVSLFKGEALPKLLPFIYAGNASNATNGNLMTGTLI
 PEKVGKIVMCDRGINARVOKVKAAGVGMILANTAAANGELVADAHLLPATVIG
 EKAGDIIRHVITDPTASISILGTGVGKPSVVAAPSPSGPNSITPNILKPLDIIA
 FGVNLIATGAGTGLASDRVFNILISGMSCPHVSGLAALLKSVHPEWSPAA
 IRSALMTTAYTKYDGPDLIDATGKPSYFDHGHVSPFTATNPLGLIDLTEDYL
 GFICALNTSPQIRSRNRYTCDPSKSYADLNYPSPFVNVNDGVGAYKRYRTVTSV
 GGAGTYSVKVTSITGVKISVEPAVLNFKANEKSYVTFTVDSKSPSGNSFGSIE
 WSDGKHVVGSPVAISMT"
 BASE COUNT 687 a 646 c 639 g 714 t

Query Match 20.7%; Score 482; DB 8; Length 2686;
 Best Local Similarity 56.2%; Pred. No. 7.5e-126;
 Matches 974; Conservative 0; Mismatches 740; Indels 18; Gaps 3;

QY 210 ACTTCTCTACCTCCTGCTCGATTGGAAGGATTTGCTGCTCAGTTGACTGAATCAGA 269
 DB 481 ATTGCTCTACACTACGAGACGCGATTTCATGGATTCTCAACACGCGTAACTCAAGAAGA 540
 QY 270 AGCGAGATCTAGATATTCACCTGAAGTTGTTGCAAGTGAGACCTGACCATGTTCTTCA 329
 DB 541 AGCGGACTCGCTCATGACTCAACCTGGTGTATCTCCGTTTACCGGAGCACCGTTACGA 600
 QY 330 GGTTCACACCACTTACTCTTACAGTTCTTCGGGACTCGACGGTTTGGAAACTCCGGTGT 389
 DB 601 GCTACACACCACTGCTACT 657
 QY 390 ATGGTCTAAATCTCGTTTGGTTCAGGACACAAATATATCGGCGTGTGTTGATCTAGGATTG 449
 DB 658 GTTCCCTGAAGCGGCTCTTACAGCAGCTGCTCGTGGAGTTCTCGATACCGGAGTTG 717
 QY 450 GCCTGAAGAGTCTAGCTTTGACGATACCGGAATGCTCTTCGATTTCACGGAAATCGAAAGG 509
 DB 718 GCCCGAGAGTAAAGCTACTCCGACGAAGGATTCGGTCTCTATCTCTCTCTCTCTCTCTCT 777

QY 510 GATTTCGAAGAAGAGAAAGTTTTCAGTTCTTCGAGCTGTAAACCGGAAGCTTAATCGGTGC 569
 DB 778 CGGATCGGAGCGCGGAACCAATTTACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGGAGC 837
 QY 570 TAGATTTCTTATCAGAGGACACCGTGTGCTGTAATTCACACAGAGGAATCACCAACATGCC 629
 DB 838 AAGATTCTTCGCTCGTGGTTTAC-----GAATCAACATGGGACCAATCGATGAATC 888
 QY 630 TCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCCGCTCAACAGT 689
 DB 889 TAAAGAAATCAAGATCTCTTAGACGACGACGCGTCAACGGAATCTCACCTCATCGACCGC 948
 QY 690 TGGTGGATCTCTCTGTTTCGATGGCGAATGTTCTTGGCAATGGAGCTGGTGTGGCTCTGTG 749
 DB 949 CGCTGGATCCGTGCTTGAAGGAGCTAGCTTATTAGGCTACGCTTCAGAAACAGCTCGTGG 1008
 QY 750 GATCGCTCTGGAGCTCACATTTGCACTTATAAAGTCTGTGTTCAATGGTGTTCACAG 809
 DB 1009 TATGGCTCCACGCGCTCGTGTGCTGTTTACAAAGTCTGTGGCTCGGTGTTGTTTCAG 1068
 QY 810 CTCTGACATTTAGCAGCTATAGATGTAGCGATTCAAGATAAAGTTCGATGTTCTTTCGCT 869
 DB 1069 CTCAGATATTTAGCTGCAATCGATAAAGCCATCGCCGATTAACGTCAATGTCTCTGTCAT 1128
 QY 870 TTCCCTTGGCGGTTTCCCTATTCTCTTGTATGATGACACCAATCGCCATTGGAAACATTCG 929
 DB 1129 GTCACCTCGCGTGGAAATGTCGGATTATTATAGAGACGGTCTTCGATCGGAGCATTCGC 1188
 QY 930 AGCCATGGAACCGCGTATATCTGTAATCTGTGAGCTGGTAAACACGCTCAATCGAAAG 989
 DB 1189 CGCCATGGAAGAGGGATTTTAGTATCTGTCAGCTGGTAAATGCTGCTCGGAGTAGTTC 1248
 QY 990 CTCTGTTGCAAAACACAGCTCTCTGGGTCTCAACCATTTGGCGCAGGACGCTTGTATCGAAG 1049
 DB 1249 TAGTTTATCAAAACGCTAGCTCCATGATGATCAACATGTTTGGTGTGGTCTTATTAGATCGTGA 1308
 QY 1050 ATTTCCCGCTGTGCTGAGATTAGCCACGGAAGAGTCTCTATGAGAGTCAATGTATGCC 1109
 DB 1309 TTTTCCCGCGCTTGGGATCTCGGTAACGGAAGAAATTCACCGAGTTCCTTGTGTTAA 1368
 QY 1110 GGGAAAGGTATAAAGATGCGCGGAGAGAGTTGAGGTGATTTACGTCACAGAGGAGA 1169
 DB 1369 AGGAGAACTTTACCGGATAAATTTGTTCCGTTTATTTACGCTGGGAATGTAGTAATGC 1428
 QY 1170 TAAAGGAAGTGAAGTTCTGTTGAGAGGTCTACTTCCAAAGAGAAATTCGAGGCAAAAT 1229
 DB 1429 TACTAATGCTAATCTCTGTATGACCGGAACCTTGTATCCCGGAGAAAGTAAAGGGGAAGAT 1488
 QY 1230 GGTGATTTGTGATCCGCGAGTCAATGGAAGATCGGAGAAAGAGAGCGGTTTAAAGAGC 1289
 DB 1489 TGTGATGTGTATAGAGGAATTAATGCTAGAGTTTCAAGAAAGGTGATGTGTTAAGCAGC 1548
 QY 1290 TGGAGGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCAAGAAAGAAATTTCTATTGA 1349
 DB 1549 AGGTGGAGTTGGAAATGATTTCTGGCTTAACACTGCGCGGAATGGTGAAGAGCTTGTTCGGA 1608
 QY 1350 CGTTCATCTCTTACCAGCTTACATTTGATTTGGTTTACACTGAGTCAGTCCCTTCTGAAGCTTA 1409
 DB 1609 TGTCTATTTGTTACCGGACCCACCGTTGGTGGAAAGACCGGTGATATAATCCGCAATTA 1668
 QY 1410 TGTTAATGCCAGGTGAACCAAGCGCGGAGTAATTTTGGTGTGTCGTTGATTTGGGAG 1469
 DB 1669 CGTGACTACTGATCTTAATCCCGCTTCGATTTCAATCTTTAGGAACAGTCTCGTGTGT 1728
 QY 1470 GTACGAGCACCGGAGTGGCTCAGTTTTCAGCTCGAGACCGAGTTTATGCAATTCCTTC 1529
 DB 1729 TAAACCATCTCCGGTTGTCGAGCGTTTAGCTCAGGTGACCCGAATTCGATTTACCCGNA 1788
 QY 1530 GATACATAAACCGGATATGATTTGCTCCGGAGTCAATATCATTTGGGCTTGGCCTCAAAA 1589
 DB 1789 TATCTTAAACCGGATCTGATCGCTCTCTGAGTAAACATCTCTCGCGCGGTGGACCGGTGC 1848
 QY 1590 TCTAGGACCAACCGGACTTCTTATGATTCAGAAAGAGTTAACTTCACTGTAATGTGAGC 1649

```

Db      1849 TGCTGACCAACCGGAGTCTCGGATTCCTCGCGCGTGAGTTCATATCATCTCTGG 1908
QY      1650 AACTTCAATGCTGTGCACATGTTAGCGGAATCACTGCTCTTATCGGCTGATACCC 1709
Db      1909 GACGTCGATGCTTTCCTCCCTACGTTAGTGGTTAGCGCGGCTTCTCAAGTCGGTGATCC 1969
QY      1710 GAACCTGCTCCAGCTGCAATCAATCCGCAATGATGACACAGCGGATTTGACGATCG 1769
Db      1969 TGAATGGAGCCCGCGGATGATGATGCGGCTTATGACACCGCTTACAAACCTACAA 2028
QY      1770 TCAAGGGAAGGATA-----AAGGATGTTAAACACCGCGGTGTTGCGATGG 1823
Db      2029 AGACGCTAAACCGTCTACTCGACATCGACACGGAAGCCTTCGACGCGGTTGATCATCGG 2088
QY      1824 AGCAGGCGATGTAATTCGCAAAAGCGGATTAACACCGGATTTGTTACACATTCACAC 1883
Db      2089 TGCAGACACGTTGTCACCAACACTGCCACTATCCAGGATCTATCAGATCTAAGCAC 2148
QY      1884 AGTGGATTACATACTTACCTCTGCACTCTTGGATTCACAAAGATCAGATATT 1935
Db      2149 GGAGGATTACTTAGGCTTCTCTGCGCATGGAATTACACATCGCGCAAAATT 2200

```

RESULT 6

```

AB007645/c
LOCUS      72698 bp      DNA      linear      PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K8K14.
ACCESSION AB007645 BA000015
VERSION    AB007645.1 GI:2564045
KEYWORDS
SOURCE     Arabidopsis thaliana (thale cress)

```

ORGANISM

```

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (sites)

```

```

REFERENCE
AUTHORS    Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.
            and Tabata,S.
TITLE       Structural analysis of Arabidopsis thaliana chromosome 5. III.
            Sequence features of the regions of 1,191,918 bp covered by
            seventeen physically assigned P1 clones
JOURNAL     DNA Res. 4 (6), 401-414 (1997)
MEDLINE     98162728
PUBMED      9501997

```

REFERENCE

```

2 (bases 1 to 72698)
Nakamura,Y.
Direct Submission
Submitted (03-OCT-1997) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan [E-mail: ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934]
Address for correspondence: kaos@kazusa.or.jp

```

COMMENT

```

For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agg_graph.cgi?c=K8K14
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://grenlil.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Rddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K3G17 and the 3' clone is K919.
Location/Qualifiers

```

```

1. 72698
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="K8K14"
/clone_lib="Mitsui TAC"
complement(join(283..653,770..1619))
note="contains similarity to unknown protein"
gene_id:K8K14.1
sp|Q10058"
/codon_start=1
/evidence=not experimental
/protein_id="BAB09014.1"
/db_xref="GI:9758428"
/translaton="MAVISLLPFPVLTTRITICNSSKSMAAKQIQISERTKRVFV
CGGVIQVCTAYFLAKGIATVLEKSAVACAAGGAGFLAFDWCGGDFVAHLAAS
FKLHSAELNGVSYGYRALTTLVTVTESKPGGGGLGPDWNGPKAKSTITVT
QTAAHPQLPFLKLSITAREKIVGVYVLEKLEVEVSGRVNSVLEGGVITDTPV
VLAMPKSKFELLSSIFRYCTKAHSIVLEKPEPAITPHALFLYRPAHGEALDP
VAMPPTGEVYICGMSQEEVPDADQVTSNPESIOVLKRVKTVSSVLYNEALVKA
EQACFLSTEDGYPVIGEIPGIKGYVGTGHSWGILNGPATGAALAEILVDVATSV
DLRSPSPSRFSKR"
5973..6890
/note="gene_id:K8K14.2
pir||S71284"
/codon_start=1
/evidence=not experimental
/product="myb-related protein, 33.3K"
/protein_id="BAB09015.1"
/db_xref="GI:9758429"
/translaton="MADRIKGPSPDEOLRLVVKYGRNWTVISKIPGRSGKSC
RLRNCNLSQVVEHREPESADETIAHAQFCNKWATLRLNGTIDVAKNHNST
LKRKCGYHRGVDGSEDEHPKRSYAGSPVVTGLYMSPTGSDYSDSTIPL
PSVLFKFPVPGAVLPLFISSSSDDPTSLSLPGAUVSESNKSHSNTINN
TTSRHNHNTVSMFSPGGFRAIGEMKSGFPGNGGFMVAVQEMIKAEVRSYMTM
CRNNGGFGVGFIDNGIMPNSQICVGRIE"
complement(join(7177..7782,7871..8755))
/note="gene_id:K8K14.3"
/codon_start=1
/evidence=not experimental
/product="cytochrome P450"
/protein_id="BAB09016.1"
/db_xref="GI:9758430"
/translaton="MGEMIDFLALMAGFIAVAVFRSKOKNLPNVPFVIGHL
HLKEPVHRSRLDLNLGIDVILGLSERAVVTASAEEFLSQQNDVFAEPLI
ATLEYMGXNLTIVSTAPYGEHRRRLRPFCAVDILSTARLDFDRDEVRAMIRKI
NVELVTGSGVRLKLOPFLYGLTYNLMMSVAGKREDEETKEVRKLREVFAGVN
YVGFPLTLKLDLGYRRAKLAGLDFKQKLVDEHRRKNGKAELEKMTIRLLS
LQSEPECYTDDIIGLVQVMLAGTDTTAVTLEWAMNLLNHPVLRKLKTELNEVS
KEGVPFRESDTGKCPYLNNTVSETLRFPAAPLLVPHASSTCEVAGFDI PRRTWLF
NAAIQRDPNVDDPETFKEPSESETHRGKFLPGIGRACPGMLAQIIVLSLGS
LIQCFWERNDVAVMSQGLTMFKSPVLVAKCSLFLDKLVJ"
join(9132..9192,9618..9763,9851..10349,10459..10523,
10630..10676,10761..10911,11066..11129,11503..11661,
11749..11870,11983..12057,12141..12194,12290..12363,
12464..12538,12619..12838)
/note="gene_id:K8K14.4
ref|NP_005638.1
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB09017.1"
/db_xref="GI:9758431"
/translaton="MSLSIVSLNVLVFRYLBQSGTHAAFTLYGAGINKSNIDGNM
VPKALLKFKVQKGLQYMEANLSEVIDDFSPFDLLISKDVRELQDMLREKK
RKRDMKEDKSKENDKREHEDGRNRAKEKREKEREKEREKEREKEREKEREK
EKIEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREK
EMVTDQTKRIAGDGTGAEPMDIVMTPTSTQSHIPNSDVRILEGTSVCAWAPS
ASLASGSGATARIWSIPGSPKAVHTGTNNALILHAKGSKNEKSKDVTLDWNG
EGTLLATGSCDGGARITWTLNGELISTLSKHGPIFSLKNKKGIDYLLTGSVRTAVV
DVRAEBWKQQFESFGPTLDDWRNNVSTATSDSNILYLCKIGETRRPARKFTTGHO

```


VNCWDPDPTGSLASCSDDSTAKIWNKIOSTFVHDLREHTKELYTIIRWSPGTGNTNP
NKQITLASFDSTVKLDAELGKMLCSFNGHREPYVSLAFSPNGYIYLAGSLDKSIH
IWSIKGIVKTYTGNGGIEFEVCWNKEGNIKACFADNSVCVLDPRM"
join(13686, .14209,14347, .14483,14567, .15444)
/note="gene_id:K8K14.5"
/codon_start=1
/evidence=not_experimental
/product="natural resistance-associated macrophage
protein"
/protein_id="BAB09018.1"
/db_xref="GI:9758432"
/translation="MSSTRDRPLASEERAYETEKVLIVIGIDEEDADYDDPQNS
PKFSKKLWLPFGFPLMSIAFLDPGNLESDLOAGATAGSLIWLAWMTAIGLLIQL
LSALGATGRHLAEKREYPTWARMLVMAEIALIGADIEVIGSAIAIKILNG
LVPLWAGVITADCFIFLENGYIRKLEAVFAILLIATMALAFVFGGPKSFDEL
LVGALVKLSRTIIVQVIGVIGIMPHVFLHSALVQSEVDPKFRFRKALRYEL
IESGALAVSFTIIVFVFAKSYFYGTETADTIGLANAGOVLDKYGFFPILYIW
AIGYLAQGSSTIGTVAGOFIMGGFLNLKWKWVRALITRSCAIIPTMIVALVPS
DSMDLEMLNVLSQVIFPFAVPIPLCLVNSQIMGSKFIQPLVQITMIVALVIA
INGLYMVDFFSGAATNLIILVIFAIAIVFVLYLISRGLEYITPWLVAASHKSPQ
RDDB"

join(16968, .17243,17316, .17432,17512, .19215)
/note="gene_id:K8K14.6
pir||T00518
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09019.1"
/db_xref="GI:9758433"
/translation="MEVSWLRVLVDNLISSVLS:SSMDLSSNPAHKYTYRGDIGKLI
KPVLENDSDAASELNLNGFEELAQYVDELREQFQSQWPLSTRIFYVILRISLASK
LRESLEVFQLLKECEQLPADLISPFEEICELKLVARDEISYITDQALKDQKGV
GPTSEVLVPAESGLRANOBILVEGVLTNMRKDELTNDTEAEYLDGLISLTQM
HEYLSDIKOALRCPVPSPDFRCSLSLEMTDPIVASGTFERFVIOKWIIMGLMW
CPKTRQALSHITLPIVTRAFPLASWCETNNVYPPDLELIHSEEPPLAVESVRASS
SENGHSLDEELRQVFSASAPGIVSEVCKTKNNNAADRSLTRSTWKPPE
ERHRHPGIIIPATVRETCSSSIETKQLIDDDSSLDTOREATRIILARNST
NRIIVARKEANSVLSLYSTDERIQADAVTCLNLSINDNKSLLAESAIVPLIHLV
LKTGVLEAKANSATIFSLSVIEYKTEIGEAGAIPLVDLLGSLGSKDAALIL
FNLISHENKTKVTEAGAVRYLVELMDPAPGVEKAVVVLANLATVREGKIAIGEGG
IPLVVEVVELSGARGKNATAALLQLCTHSPKFCNNVIREGVIPPLVALYKSGTARGK
EKVLPLPPLCLVNVNS"

complement(20385, .21293)
/note="gene_id:K8K14.7
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09020.1"
/db_xref="GI:9758434"
/translation="WAMTCTEDDOKTLILSHRYOEDDDVEEBEERALSICDLPNK
GELSLVKEEEDERDSPEFCIGSFRAGSDSCPAPEMSTADELPFKGILPLRHVS
SLDAGLNPTLIRTSSEVPRRTGIIRSDRKIKNFIDYSQSPQPIRISSTMTAR
VNSIRNPKSSSIWPLRLGLVTRTEIELIRTAGNKLUSVRNSSCSSTISSSKKIG
SGESRNRNRSFMFSDCKSVSTETKMAPVKIKVSSGETEERKMMKRTAKKEKS
AMARXRTPEWLSQGVFVVDHGRRLV"

complement(24056, .26329)
/note="gb|AAC18851.1
gene_id:K8K14.8"
/codon_start=1
/evidence=not_experimental
/product="cucumisin-like serine protease"
/protein_id="BAB09021.1"
/db_xref="GI:9758435"
/translation="MSSSFLSSTAFFLLLCGFCHVSSSSDDQGTYYIVHAKSMPSS
FDLHNSNYSLSRSDSAELLYYENAIHGFSRLTQBEADSLMTQGVISVLPERH
YELHTRPTFLGIDETADLPFPAQSYSDVVGVLDTPGWPEKSDRQFGFIPSS
WKGGEAGTNTFASLNKRLICARPPARGYESTWGPIDESKPSRSDDDGHHGTNS
TAAGSVGASLILGYASTARGMAPARVAVYKVCWLGCGPSSDILAIDKADNVN
VLMSIGGSGSDYRDVAIGAFAFAMERGIILVSSAGNAGPSSSLSNVAPWITTVGA
GTLDRDFPALAILNGKNPTGVSLFKGALPKLLPFIYAGNASNATNGNLNMTGLI
PKVKVKGITVMDCRNQRQVQKVVAAAGGMILANTPAANGEELVADAHLLPATVVG
EKAGDILIRHYVTDPNPTASISILGTVVGVKPSVVAAFSSRGSNTIPNLIKPLDIA
PGVNIILAAWTGAGTGLASDRRVEPNIISGTSNCPHVSLAALLKSVHPWSPAA

CDS

CDS

CDS

CDS

IRSLMTTAYKYKDGKPLLDIATKPSPTPFHDGAGHVSPTTATNPGLIYLDTTEDYL
GFLCALNYSQIIRSVRRNYTCDPSKSYSVADLNYPSPFAVNDVGAYKAYKTRVTSV
GGAGTYSVKVTSSETVTKIISVEPAVLNFKEANESKSYVTFTVDSKSPSGNSRGSIE
WSDGKHVGVSPVAISW"
complement(join(29531, .29593,29772, .29953,30043, .30078,
30480, .30575,30665, .30742,30830, .30982,31054, .31329))
/note="gb|AAC18972.1

Query Match 20.7%; Score 482; DB 8; Length 75698;
Best Local Similarity 56.2%; Pred. No. 8.9e-126; Indels 18; Gaps 3;
Matches 974; Conservative 0; Mismatches 740;

QY 210 ACTTCTCTACTCTCTGCTCTGCGATTGAAGGATTTCTGCTCAGTTGACTGATCAGA 269

Db 26138 ATTGCTCTACATTACGAGACGCGATTCTGGAATCTCAACACGGCTAACTCAAGAAGA 26079

QY 270 AGCCGAGATCTAGATATTCACTGGAAGTTGTGTGACGTAGACACTGACCATGTTCTTCA 329

Db 26078 AGCCGACTCGCTCATGACTCAACCTGGTGTATCTCCGTTTTACCGGAGCACCGTTACGA 26019

QY 330 GGTTCAAACCACTTACTCTTACAACTTCTTGGGACTCGACGGTTTGGAAATCCGGTGT 389

Db 26018 GCTACACACCACTCGTACTCTCTCTCTCGGTCTCGACG---AACACACCGCAGATCT 25962

QY 390 ATGGTCTAAATCTCGGTTTGGTCAAGGCACAAATATCGCGCTGCTTGATATCTGGAGTTTG 449

Db 25961 GTTCCCTGAAGCCGCTCTTACAGCAGCTCGTGTCTGAGATTCTCGATACCGAGTTTG 25902

QY 450 GCTCGAAAGTCTTAGCTTTGACGATACCGGAATGCTTTCGATTCACCGAAATGGAAGG 509

Db 25901 GCCCGAGAGTAAAGACTACTCCGACGAAGGATTCGGTCTTATTCCTCTCTCGAAAGG 25842

QY 510 GATTTCGCAAGAGAGAAAGTTTCACTTCTTGGAGCTGTAAACCGGAGCTTAATCGGTGC 569

Db 25841 CGGATCGAGGCGCGAAACCAATTTTCCGCTTCTCTCTGTAACCGTAACATCGGAGC 25782

QY 570 TAGATTCITCATCAGAGCACCGTGTCTGTAATTCACGAGGAGTACCAACATCGC 629

Db 25781 AAGATTTCTCGTCTGGTTAC-----GAATCAACGATGGACCAATCGATGATC 25731

QY 630 TCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACCGCTCAACAGT 689

Db 25730 TAAAGATCAAGATCTCTTAGAGACGACGCGTCAAGGACTCACCTCATCGACGC 25671

QY 690 TGTGGATCCTCTGTTTCGATCGCGAATGTTCTTGGCAATGAGCTGTGTGGCTCGTGG 749

Db 25670 CGCTGGATCCGCTGTGAAGGAGCTAGCTTATTAGGCTACGCTTACAGAAACAGCTCGTG 25611

QY 750 GATGGCTCTGAGCTCACTTGCAGTCTATPAAAGTCTGTGGTTCAATGGTTGTACAG 809

Db 25610 TATGGCTCCACGCGCTCGTGTGCTGTTTACAAAGTCTGTGGCTCGTGGTTGTTT 25551

QY 810 CTCTGACATCTAGCAGCTATAGATGTAGCGATTCAAGATAAGTCGATGTTCTTCGCT 869

Db 25550 CTCAGATATTTAGTGTCAATCGATAAGCCATCCCGGATACGTCATGCTCTCGAT 25491

QY 870 TTCCCTTGGCGGTTTCCCTATTCTTCTGTATGATGACACAAATCGCCATTGGAAACATCCG 929

Db 25490 GTTACTCGCGGTGGAAATGTCGGATTATTATAGACGGGTGTGCGATCGGAGCATTCG 25431

QY 930 AGCCATGGAACCGGTTATATCTGTAATCTGTGCACTGGTAAACACGCTTCAATCGAAG 989

Db 25430 CGCCATGGAAGAGGGGATTTAGTATCTTGTCTCAGCTGGTAAATGCTGGTCCGAGTAGTTC 25371

QY 990 CTCTGTTGAAACACAGCTCTTGGTCTCAACCAATTGGCGCAGGACCGCTTGATCGAAG 1049

Db 25370 TAGTTTATCAACAGTACGTCCATGATCAAACTGTGGTGTGGTACTTTAGATCGTGA 25311

QY 1050 ATTTCCCGCTGTGGTCAGATTAGCCAAACCGAAAGCTTCTCTATGAGAGCTCAATCTATCC 1109

Db 25310 TTTTCCGGCGCTTGGCATTTCTCGTAAACGGCAAGAAATTTACCGGAGTTTCTTTGTTAA 25251

QY 1110 GGGAAAAGGTATAAAGATCCGGGAGAGAGGTTGAGGTGATTTACGTCACAGGAGGAGA 1169

CDS

CDS

CDS

CDS

Db	25250	AGGAGAGCTTTACCGGATAAATTTGTCGGTATTTATTTACGCTGGGAATGCTAGTAATGC	25191
Qy	1170	TAAAGGAAGTGAGTCTTGTTTGAGAGGGTCACTTCCAAGAGAGAAATCCGAGGCAAAAT	1229
Db	25190	TACTAATGGTAATCTCTGTATGACCGGAACTTTTGATCCGGAGAAAGTAAGGGGAAGAT	25131
Qy	1230	GGTGAATTTGATCGCGGAGTCAATGAAGATCGGAGAAAGAGAAAGCGGTTAAAGAAC	1289
Db	25130	TGTGATGTGTAGAGAGAAATTAATCTAGATTCAGAAAGGTGATGTGGTTAAAGCAGC	25071
Qy	1290	TGGAGGATTTGCAATGATCTTACCAATACAGAGATCAACCAAGAGAAAGATCTTATTCGA	1349
Db	25070	AGTGGAGTTTGGATGATCTTGGCTAACACTCGCGGCAATGGTGAAGAGCTTGTTCGGGA	25011
Qy	1350	CGTTCACTCTTACAGACTCAATGATTTGTTTACACTGAGTCACTCTCTCTCAAGGCTTTA	1409
Db	25010	TGCTCATTTGTTACCGCGGACCCGTTGTTGTAAGAACCGGATGATTAATCCGGCATTA	24951
Qy	1410	TGTTAATGCCACGGTGAACCAAGCGCGGATTAATTTTGGTGGTACGGTGAATGGGAG	1469
Db	24950	CGTGACTACTGATCCTAATCCACCGCTTCGATTTCAATCTTAGGAACAGATCGTCGGTGT	24891
Qy	1470	GTCAGAGACCGGAGTGGTCTGATTTTTCAGCTCGAGGACCGAGTTTACCCAAATCCTTC	1529
Db	24890	TAAACCATCTCCGGTTTTCGACGGTTTAGCTCACGTGGACCGAATTCGATTAACCGGAA	24831
Qy	1530	GATACATAAACCGGATATGCTCCGGAGTCAATATCATTTCCGCTTGGGCTCAAAA	1589
Db	24830	TATTCCTTAAACCGGATCTGATCGCTCTGAGTAAACATCTCGCGCGTGGACCGGTGC	24771
Qy	1590	TCTAGGACCAACCGGACTTCTTATGATTTCAAGAAAGATTAACCTCACTGTAATGTGAG	1649
Db	24770	TGCTGGACCAACCGGACTCGCTTCCGATTTCTCGCGCGTGGAGTTCAATATCATCTCTGG	24711
Qy	1650	AACCTCAATGCTTTGTCACATGTTAGGGAATCACTGCTCTTATCCGGTCTGCATACC	1709
Db	24710	GACGTCGATGCTTTGCCCTCACGTTAGTGGTTAGCGCGCTTCTCAAGTCGGTGCATCC	24651
Qy	1710	GAACGCTGCTCAGCTGCAATCAATCCGATTTGATGATGATGATGATGATGATGATGATG	1769
Db	24650	TGAATGGAGCCGCGCGGATTTAGTACGGCGCTTATGACCAACCGCTTACAAAACCTACAA	24591
Qy	1770	TCAAGGAAAGCGATA-----AAGGATGGTAACAAACCAACCGCGGTGTTTGGATTTG	1823
Db	24590	AGACGGTAACCGTTACTCGACATCGCGACAGGAGAGCCTTCGACGCGTTCGATCACGG	24531
Qy	1824	AGCAGGCGATGTGAATCCGAAAGGCGATAAACCCGGGATGGTTTACAACTTCAACC	1883
Db	24530	TGCAGGACACGTGTCAACCAACACTGCCACTAATCCAGGACTCATCTACGATCTAACGAC	24471
Qy	1884	AGTGATTAATACATTAATCTACCTGCACTCTTGGATTCACAAGATCAGATATT	1935
Db	24470	GGAGGATTAATTAAGGCTTCTCTCGCATTAATTAACATACATACGCGCGCAATTT	24419

RESULT 7
AX506370
LOCUS 2265 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 1065 from Patent WO0216655.
ACCESSION AX506370
VERSION AX506370.1 GI:23387607
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
AUTHORS Stress-regulated genes of plants, transgenic plants containing
TITLE same, and methods of use
JOURNAL Patent: WO 0216655-A 1065 28-FEB-2002;

FEATURES	source	Location/Qualifiers	The Scripps Research Institute (US) ; Syngenta Participations AG (CH)
BASE COUNT	519 a	-510 c 542 g 694 t	
ORIGIN			
Qy	380	ACTCGGTGATGCTTAATCTCGGTTTGGTCAAGGCAAAATATATCGCGTGGTGTGATA	439
Db	344	AATTCGGTGTTCACGATCTCGGTTCTTCTTCAACGGCGTTATCATCGGAGTTTTAGATA	403
Qy	440	CTGGAGTTTGGCCCTGAAGTCTTAGCTTTCAGATACCGGAATGCCCTTCGATTCACCGGA	499
Db	404	CTGGGTTTGGCCCTGAATCTTAGAAGCTTCGATGATCTGATATGCTTGAGATTCCTCTTA	463
Qy	500	AATGGAAGGGAATTTGCCAAGAGGAGAAAGTTTCAGTTCTTTCGAGCTGTAAACCGGAAGC	559
Db	464	AATGGAAGGAGAAATGTGAATCTGTTCCGATTTTCGATTCGAAGTTGTGTAAACAAGAAGC	523
Qy	560	TAACTGGTGTAGATTTCTTCATCAGAGACACCGTGTGCTAATTCACAGAGAGAAATCAC	619
Db	524	TATTCGGAGCTAGAAGCTTCTCCAAAGGATTTCAATTCGGCTTCGTGGTGGTTTTTCGA	583
Qy	620	CARACATCCCTCGTGAATACATTTCCGCAAGAGATTCACCGGACACGGGACTCACACCG	679
Db	584	GTAA-----CGGTGAATCTGTTCTCTCTCGTGAATTTGACGGACATGGAACATACTT	637
Qy	680	CCTCAACAGTTTGGTGGATCTCTGTTTCGATGGGGAATGTTCTTGGCAATCGAGCTGGTG	739
Db	638	CMACTACCGCGCGGATCCCGCTTAGAAGCTAGCTTCTCGTTTACCGCCGGTA	697
Qy	740	TGGCTCGTGGAGTCTCCTGGAGCTACATTCAGCTATATAAGTCTGTGTGGTTCATG	799
Db	698	CGGCGAGAGGTATGCCCACCTCGTCTGTTGCTTACTTATAAAGTTTGTGGAGTACTG	757
Qy	800	GTGTTTACAGCTCTGACATTTCTAGCAGCTATAGATGACGATTCAGATTAAGTCGATG	859
Db	758	GTGTTTGGATCTGATATCTAGCTGTATGGATCGAGCTATCTTGATGGTGTGATG	817
Qy	860	TTCTTTTCGTTTCCCTTCGGGTTTCCCTATTTCTTCTTATGATGACACAATCGCATTG	919
Db	818	TGCTTTCTGTTATCTCTTGGTGGTGTCTGCTCCGATTTATCGGATACGATTCGATG	877
Qy	920	GAACATTCGAGCCATGGAACCGGATATATCTGTAATCTGTCAGCTGGTAAACCGGTC	979
Db	878	GAGCGTTTTCGGCTATGGAGAGAGGTGTTTGTGCTTCTGCTGTTTCTGTTAATAGTGTC	937
Qy	980	CAATCGAAAGCTCTGTGCAACACAGCTCTTGGGCTCTCAACATTTGGCGGACGACGC	1039
Db	938	CTACTAGAGCTTCTGTGCGCAATGTGCTCTTGGGTTATGACTGTTGGTGTGCTACTT	997
Qy	1040	TTGATCGAGATTTCCCGCTGGTTCAGATAGCCCAACGGAAAGCTTCTCTATCGAGAGT	1099
Db	998	TAGATAGAGATTTCCCGGCTTTTGGCAATCTCGGTAAACGGGAACGACTTACCGGTGTT	1057
Qy	1100	CATTGTATCCGGAAAAGGTATAAAGATGCCGGAGAGAGGTTGAGGTGATTTACGTCA	1159
Db	1058	CGCTGTATAGCGGTGTAGAAATGGGGA-----CGAAGCGCTTGAATTTGTTTATAATA	1111
Qy	1160	CAGGAGGAGATAAAGAAAGTGTGCTGTTTGAAGGGTCACTTCCAGAGAGAAATATCC	1219
Db	1112	AAGGGAATAGTAGTTCGAGTAATCTTGTGTTACCTCGGTTGCTGATTCAGTATTGTTTC	1171
Qy	1220	GAGGCAAAATGTTGATTTGTGATCCGGAGTCAATGGAAGATCGGAGAAAGGAGAGCGG	1279
Db	1172	GTGGGAAGATTTGTTGTTTGTGATAGAGGTGTTTAATGCTAGATTTGAGAAAGGAGCTGTG	1231

QY	1280	TTAAAGAAGCTGGAGGATTGCAATGATCTTAGCCAAATACAGAGATCAACCAGAAGAAG	1339
DB	1232	TTAGAGATGCTGGTGGTTTAGGGATGATAATGCGGAATACTGCTGCGAGTGGAGAGGAC	1291
QY	1340	ATTCTATTAGGTTTCATCTCTTACCAGCTACATTGATTGGTTACATGAGTCAGTCCCTTC	1399
DB	1292	TTTGGCGGATAGTCAATTTGCTTTCCGCGATCGCTGTAGGGAAGAAGACTGGTGATTTAC	1351
QY	1400	TGAAGCTTATGTTAATGCCACGGTGAACCAAGGCGGAGATAATTTTGGTGGTACGG	1459
DB	1352	TTAGGAGTATGTTAAGTCAGATCTTAACCAACCGCTCTTCTTGTTTTTAAGGAACGG	1411
QY	1460	TGATTGGGAGGTCACGAGCACCGAGGTGCTCAGTTTTTCAGTCTCAGGACCGAGTTTAG	1519
DB	1412	TTCTTGAGGTTAAGCCGCTCTCCTGTGGTTGCTTGCTTTAGCTCGAGAGGTCTTAATCTG	1471
QY	1520	CCAACTCTTCGATACATAAACCAGGATATGATTGCTCCGGAGTCAATAATCATTTGCGGCTT	1579
DB	1472	TTACTCTCGAAATCTTGAAGCCTGATTTATTGGTCTGGAGTTAATATTTTGGCTGSGTT	1531
QY	1580	GGCTCAAAATCTAGGACCAACCGGACTTCCTTATGATTCAAGAAGAGTTAACTTCACTG	1639
DB	1532	GGTCTGAGCTATTGGTCTCTACTGGTCTTGACAGGACTCTAGGAGACTCAGTTCAACA	1591
QY	1640	TAATGTCCAGGAACCTCAATGCTTGTGCCATGTTAGCGGAATCACTGCTCTTTACCGGT	1699
DB	1592	TCATGTCAGGTACGTCAATGCTCATGCGCCACACATCAGTGGTTTAGCGGTCCTTTTGAAG	1651
QY	1700	CTGATACCCGAACCTGCTCTCCAGCTGCATCAAAATCCGCAATTGATGACAAACAGCGGATT	1759
DB	1652	CAGCTACCCCTGAGTGGAGTCCGAGTCTCTATCAAAATCAGCTCTCATGACTACAGCTTACG	1711
QY	1760	TGTACGATCGTCAAGGGAAGCGATAAGATGGT-----AACAAACACAGCCGGTGTGT	1813
DB	1712	TTCTTTGAACCAACAGCTCCTCTCCATGATGCTGCAGACACAGCCTATCTAACCCAT	1771
QY	1814	TTGCGATTGGAGCAGGCGATGTGAATCCGCAAAAGGCGATAAACCCGGGATTGGTTTACA	1873
DB	1772	ATGCTCACGGCTCGGGCCATGTAGATCCCAAAAGGCTCTCTCACCAAGGTCCTTGTCTACG	1831
QY	1874	ACATTCACAGTGGATTACATAACTTACCTCTGCATCTTGGATTTCAGAGATCAGATA	1933
DB	1832	ACATCTCAACCGAGGAATACATCAGGTTTTTGTCTCTCTAGACTACACAGTCEATCACA	1891
QY	1934	TTTTAGCAATCACTCATAGAAGCGTAGCTGCAATGGAATATTCGGGAAAAACCCGGGTT	1993
DB	1892	TTGTTGGCAT-----TGTGAAGCGACTAGCGTTAACTGCTCGAAGAGTTCTCAGATC	1945
QY	1994	TTAGTCTCAATTAACCGTTCGATAGCCGTGATTTTTCAAAACGTGGCAAGACTACGGAGATGA	2053
DB	1946	CTGCTCAGCTCAACTACCCAAAGTTTCTCGGTTTTTGGGGGTAAAGAGTTGTGCGGT	2005
QY	2054	TCACAAGCGTGTCACTAAAGCTGGAGTCTTAATCTCGATATACTAGTGAATGTCAAGG	2113
DB	2006	ACACTCGGGAAGTAACAAATGTTGGTGCAGCAACCTCGGTTTACAAGTGCACGTTAATG	2065
QY	2114	CTCCAGAGGGGATCAAAAGTTATTGCTCAATCCTAAGAGACTTGTCTTCAAAACAGCTGGATC	2173
DB	2066	GAGCTCCTAGTGTGGAAATCTCTGTTTAAACCATCGAAACCTTTCGTTTTAAAGCGTGGAG	2125
QY	2174	AGACGCTGAGCTATAGAGTATGGTTTTGTATTGAAGAAAGAAAAACAGAGGAGGGAAGTGG	2233
DB	2126	AGAAGAAGAGGTACACAGTCACGTTTGTAGCAAGAAGGAGTGTGATGACGAACAAGG	2185
QY	2234	CTAGCTTTG	2242
DB	2186	CTGAGTTTG	2194

RESULT 8
AY142613
LOCUS

DEFINITION	ACCESSION	REFERENCE
VERSION	VERSION	AUTHORS
KEYWORDS	KEYWORDS	
SOURCE	SOURCE	
ORGANISM	ORGANISM	

Arabidopsis thaliana putative subtilisin serine protease
 (A2905920) mRNA, complete cds.
 AY142613
 AY142613.1 GI:23296837
 FLI CDNA.
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Arabidopsis thaliana
 Eukaryota, Viridiplantae; Streptophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi-
 s 1 (bases 1 to 2296)
 Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Davis, R.W., Ecker, J.R., and Theodorakis, A. 2002. Arabidopsis Open Reading Frame (ORF) Clones Unpublished
 2 (bases 1 to 2296)
 Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, B.K., Wong, C., Wu, H.C., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theodorakis, A.
 Direct Submission
 Submitted (16-AUG-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
 The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

COMMENT

The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shirozaki, K.

The Salk, Stanford, EGEC (SP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RFLP cDNAs: Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.C., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W., Ecker, J.R. and Theologos, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

FEATURES
source

```

submitted to GenBank.
Location/Qualifiers
1. .2296
   /organism="Arabidopsis thaliana"
   /mol_type="mRNA"
   /db_xref="taxon:3702"
   /chromosome="2"
   /clone="U09230"
   /note="This clone is in pUNI 51.
ecotype: Columbia"
1. .2296

```

gene

COPIES: CONTINUED

1. 2296

CDS

1. 2265

```
/evidence=experimental  
/condn_start=1
```

```

/evidence=experimental
/product="putative subtilisin serine protease"

```

```
/protein id="AAN13182.1"
```

```
/db xref="GI:23296838"
```

translation="WASSSSSSSSIIITFLLLHTTAKTIIIRVNSDXPESF
THHDWTSQJNSESLLIYTTISFGFSAYLDSFEADSLSSNSIIIDPEITGTE
TTTRPFLGNSEFVHDLCSSNGVIIGVLDTCWVPEBSFDDTDMPEI PSXWKGE
ESGSDSKLCKNKLIGARFSKGFQMASGGGCSKESVSFPRDVCCHGHTHTSTAA

SAVNASFLIYAAGTAGMATRARVATYKVCWSTCGFSGDILAAVDRAILDGVVLSL
SLGGSPAYRDTIALGAFSAMBGRVFCVSGNSGPTASVANTVPMVMTVGATGLD
RDPAPANLYNGRLGALSILYGVGMGTPLLELVNKGNSSSNLCILPGLDSSITGLD
KIVVCRGVNARVEKAUVDAGGLKIMANTAAAGEELVADSHLLPALAVGKTKGDL
LREYVSKDPTALLVFKGTLDVDFKPSVVAFAISRGENTVPEILKEDVIGPGVAIL
AGWSDAIGPTDKDDBRRTOFNIMSGTSMSCPHI SLGAGLLKAAHPWSPSAIKALIM
TTYAVLIDNTNAPLHDAADSLNPNYAHGSHVDPPQKALSPGLVYDITSEYIRFLCSL
DYTVDHIVATVKRPSVNCCKFSDPQLNYPSPFVLFPGKRVYRTREVTVNGAASSV
KYTVNGAPBSVGI SVKPSKLSFKSVGEKKRYTVTVFSKKGVSMTNKABFSGITWNPQ
HEVRSVPVBSNRP"
2266..2296
/gene="Ac2g05920"

3'UTR
BASE COUNT 532 a 515 c 546 g 703 t
ORIGIN

Query Match 20.78; Score 481.8; DB 8; Length 2296;
Best Local Similarity 55.3; Pred. No. 8.5e-126;
Matches 1033; Conservative 0; Mismatches 162; Indels 24; Gaps 4;

QY	380	ACTCCGGTGTATGGTCTAAATCTCGGTTTGGTCAAGGCACAAATTATCGCGCGTCTTGATA	439
Db	344	AATTCGGTGTTCAGATCTCGGTCTCTTCTTAACGGCGTTATCATCGGAGTITTAGATA	403
QY	440	CTGGAGTTTGGCCTGAAAGCTTAGCTTTTGACGATACCGGAATGCCCTTCGATTCACGGA	499
Db	404	CTGGCGTTTGGCCTGAATCTAGAAGCTTCGATGATCTGATATGCCCTGAGATTCCTCTTA	463
QY	500	AATCGAAAGGATTTGCGCAGAGAGGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGC	559
Db	464	AATGGAAGAGGAATGTGAATCTGTTTCCGATTTCCGATTTCCGATTTGTGTAAACAAGAAGC	523
QY	560	TAATCGGTGTAGATTCTTCATCAGAGGACACCGTGTGCGTAAATTCACAGAGGAATCAC	619
Db	524	TTATCGGAGCTAGAAGCTTCTCCAAAGATTTCAATGGCTTCTGTTGGTGGTGGTTTCGA	583
QY	620	CAAAACATGCTCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTTCACACCG	679
Db	584	GTAA-----CGGTGAATCTGTTTCTCTCGTGTATGTTGACGGACATGGAACACATATT	637
QY	680	CCTCAACAGTTGGTGGATCTCTCTGTTTCGATGCGGAATGTTCTTGGCAATGGAGCTGGTG	739
Db	638	CAACTACCGCGCGGGATCCGCGCTTAGAAACGCTAGCTTCTCGGTTACGCGCGCGGTA	697
QY	740	TGGCTCGTGGATGGCTCTGGAGCTCACATTGGAGTCTATAAGTCTGTTGTTCTCAATG	799
Db	698	CGGCGAGAGGTATGCCACTCGTCTCGTGTGCTACTTATAAGTTTGTGGAGTACTG	757
QY	800	GTTGTTACAGCTCTGACATTTAGCAGCTATAGATGTAGCGATTCAAGATAAAAGTCTGATG	859
Db	758	GTTGTTTGGATCTGATATCTAGTCTGCTATGGAICGAGCTATCTTGTATGATGGTGGTGGATG	817
QY	860	TTCTTTCCGCTTTCCGCTGGCGTTTCCCTATTCCTTTTGTATGATGACACAAATCGCCATTG	919
Db	818	TGCTTTGCTTATCTCTTGGTGGTGGTCTGCTGCTCGCTATTATCGCGATACGATTGGGATTG	877
QY	920	GAAACATTCGGAGCCATGGAACCGGTATATCTGTAATCTGTGAGCTGGTGTAAACAACGGTC	979
Db	878	GAGCGTTTTCGGCTATGGAGAGAGGTGTTTGTGTCTTGTGCTTCTGCTGGTAATAGTGGTC	937
QY	980	CAATCGAAGCTCTGTTCCAAACACAGCTCTCTGGTCTCAACCATTTGGCGCAGGCACGC	1039
Db	938	CTACTAGAGCTTCTGTTCCCAATGTTGCTCTCTGGTGTATGACTGTTGGTGGTCTGGTACTT	997
QY	1040	TTGATCGAAGATTTCCCGCTGTGGTTCAGATTAGCCAAACGGAAAGCTTCTCTATGGAGAGT	1099
Db	998	TAGATAGAGATTTTCGGCTTTTGGGATCTCGGTAAACGGGAAACGACTTACCGGTGTTT	1057
QY	1100	CATTGTATCCGGGAAAGGTATAAAGAAATGCCGGAGAGAGGTTGAGGTGATTTACGTCA	1159
Db	1058	CGCTGTATAGCGGTGTAGGAATGGGA-----CGAAGCCGCTTGAATTTGGTTTATAATA	1111
QY	1160	CAGGAGGAGATAAAGGAGTAGTCTGTTTGGAGGGGTCACTTCCAGAGAGAAATCC	1219

Db	1112	AAGGGAATAGTAGTTCAGTAATCTTTTGTACCTGGTTCGCTTGATTTCGAGTATTGTTT	1171
QY	1220	GAGCAAAATAGTGAATTTGTGATCGCGAGTCAATGAAGATCGGAGAAAGAGAGACGG	1279
Db	1172	GTGGGAAGATTGTTGTTTGTGATAGAGTCTTAATGCTAGAGTTGAGAAAGAGAGCTGTGG	1231
QY	1280	TTAAAGAAAGCTGGAGGAGTTGCATGATCTTAGCCATACAGAGATCAACCAAGAGAGAAG	1339
Db	1232	TTAGAGATGCTGTGGTGTAGGAGATGATAATGCGAATCTGCTGCGAGTGGAGAGAGC	1291
QY	1340	ATTCTATTGAGCTTCACTCTTACCAGCTACATTGATTGGTTACACTGAGTCACTGCTTC	1399
Db	1292	TTGTGGCGGATAGTCACTTTGCTTCCCGGATCGCTGTAGGGAAGAGACTGGTGTATTAC	1351
QY	1400	TGAAGGCTTATGTTTAATGCCACCGTGAAACAAAGGCGCGGATAATTTTGGTGGTACGG	1459
Db	1352	TTAGGAGTATGTTTAAGTCAGATTCTAAACCAACCGCTCTTCTGTTGTTTTTAAAGGAACG	1411
QY	1460	TGATTGGAGGTACAGAGCACCGAGGTGGCTAGTTTTTTCAGCTCGAGGACCGAGTTTAC	1519
Db	1412	TTCTTTGAGCTTTAAGCCGCTCTCTGTGTTGCTGTTTGTCTGAGAGGTCCTTAATACTG	1471
QY	1520	CCATCTCTTCGATCTAAACCGGATATGATTGCTCCGGAGTCAATATCATTCATTCGGCTT	1579
Db	1472	TTACTCTCGAAATCTTGAAGCTGATGTTATTTGCTCTGAGTTAAATTTTGGCTGGTT	1531
QY	1580	GGCTTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAAGAGAGTTTAACTTCACTG	1639
Db	1532	GGTCTGACGCTATTGGTCTTACTGCTTGAACAGGACTCTAGGAGACTCAGTTCAACA	1591
QY	1640	TAATGTCAAGAACTTCAATGCTTGTCCACATGTTAGCGGAATCAGTCTCTTATCCGGT	1699
Db	1592	TCATGTCAAGTACGTCATGTCATGCCACACATCAGTGGTGTTCGCGGCTCTTTTGAAG	1651
QY	1700	CTGCATACCCGAACTGGTCTCCAGTCCAAATCGCATTCGATTCATGACACAGCGATT	1759
Db	1652	CAGCTCACCTTGAGTGGAGTCCGAGTCTATCAATCAGCTCTCATGACTACAGCTTACG	1711
QY	1760	TGTAAGTCTCAAGGGAAAGCGATAAGAGTGT-----AACAAACCGAGCGGTGTGT	1813
Db	1712	TTCTTTGACAAACACCAACCGCTCTCTCCATGATGCTGACAGACAAACAGCTATCTAACCCAT	1771
QY	1814	TTGCGATTGGAGCAGGCGATGTGAATCCGAAAAGGCGATAAACCGCGGATGGTTTACA	1873
Db	1772	ATGCTCAAGCTCGGCTCGGCGCATGTAGATCCCAAAAGGCTCTCTCACCAGGTCTTGTCTACG	1831
QY	1874	ACATTCACACAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1933
Db	1832	ACATCTCAACCGAGGAATACATCAGGTTTTTGTGCTCTCTAGACTACACGTCGATCACA	1891
QY	1934	TTTTAGCATCACTCATTAAGACGTGAGCTCAATGGAATTTGCGGAAAAACCGCGGTT	1993
Db	1892	TTGTTTGGAT-----TGTGAAGCGACCTAGCGTTAACTGCTCGAAGAGTTCTCAGATC	1945
QY	1994	TTAGTCTCAATTAACCGCTCGATAGCGGTGATTTTCAACCGTGGCAAGACTACGGAGTGA	2053
Db	1946	CTGGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT	2005
QY	2054	TCACAGGCGTGTCTACTAAGCTTGGAGTCTTAATCGATATCTCAGTGAATGTCAAGG	2113
Db	2006	ACACTCGGAGTAACAAATGTTGGTCAGCAAGCTCGGTTTTTACAAAGTACGGTAAATG	2065
QY	2114	CTCCAGAGGGGATCAAGTATTGTCATCTTAAGAGACTTGTGTTCAACACAGCTGGATC	2173
Db	2066	GAGCTCTAGTCTCGGAATCTCTGTTAAACCATCCAAACTTTCGTTTTTAAAGCGTGGAG	2125
QY	2174	AGACCTGAGCTATAGAGTATGTTTGTATTGAAGAGAAAAACAGAGAGGAGGAGGTGG	2233
Db	2126	AGAGAGAGAGGTACACGTACGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	2185
QY	2234	CTAGCTTTTG 2242	
Db	2186	CTGAGTTTG 2194	

```

RESULT 9
LOCUS AY035090
DEFINITION Arabidopsis thaliana putative subtilisin serine protease
ACCESSION AY035090
VERSION AY035090.1
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (chale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 2541)
Yamada, K., Liu, S. X., Sakano, H., Pham, P. K., Banh, J., Chung, M. K.,
Goldsmith, A. D., Lee, J. M., Quach, H. L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R. W., Ecker, J. R. and Theologis, A.
Arabidopsis Full Length cDNA Clones
Unpublished
2 (bases 1 to 2541)
Yamada, K., Liu, S. X., Sakano, H., Pham, P. K., Banh, J., Chung, M. K.,
Goldsmith, A. D., Lee, J. M., Quach, H. L., Tang, C. C., Toriumi, M., Yu, G.,
Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,
Kim, C., Koesema, B., Lam, B., Lin, J., Meyers, M. C., Miranda, M.,
Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T., Satou, M., Seki, M.,
Shinn, P., Southwick, A., Shinozaki, K., Davis, R. W., Ecker, J. R. and
Theologis, A.
Direct Submission
Submitted (03-MAY-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN
Arabidopsis Full-Length cDNA'); Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.
The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAPL cDNAs: Yamada, K., Liu, S. X.,
Sakano, H., Pham, P. K., Banh, J., Chung, M. K., Goldsmith, A. D.,
Lee, J. M., Quach, H. L., Tang, C. C., Toriumi, M., Yu, G., Bowser, L.,
Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, B.,
Lam, B., Lin, J., Meyers, M. C., Miranda, M., Nguyen, M., Palm, C. J.,
Shinn, P., Southwick, A., Davis, R. W., Ecker, J. R. and Theologis, A.
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
contributed equally to this work as PIs.
Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.
FEATURES
Location/Qualifiers
1..2541
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF109-15-A05 (R09230)"
/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert.
ecotype: Columbia"
1..2541
/gene="At2g05920"
1..22
/gene="At2g05920"
misc_difference 2
/gene="At2g05920"
/note="compared to genomic sequence"
/replace="c"
23..2287
/gene="At2g05920"
/codon_start=1
/evidence=experimental
/product="putative subtilisin serine protease"
/protein_id="AAK595.1"
/db_xref="GI:14334834"
/translation="MASSSSSSSSIIITITFLLLHTTAKKTYIIRVNHDKPSFL
THDWYTSQNLSESLLYTTSFHFAYILDSTADSLSSNSILDI FEDPLYTLH
TTRTFELCLNSEFGVHDLGSSNGVIGLDTGVMPESRSPDDTMDPEI FSKWKEG
ESGSDFKLCKNKLIGASFSKGFOMASGGGFSKRESVSPRDVDGHTGHTTAAAG
SAVRNASFLYAGTARGMATRVARVATKVCWSTCCFSDDLAMDRAILLDQVDVLSL
SUGGSAPYRTDITIAFASAMERGVFVSCSAGNSGPTASVANVAVMTVVGAGTLD
RDFPANGKRLTGVSLYSGVMGTKPLELVYKNKGNSSNLCIPGSLDSTVIRG
KIVCDRGVNAVERKGVADAGLGMIMTAAASBEELVADSHLLPAI AVGKTKDGL
LREYVKSCKPTALLAFKGTLDVVKPSVFAAFSSRGNPTVPELTKDVTGPGVNIIL
AOWSADIPTGLDKDSRTQFNMSGTSVCPHISGLAGLLKAAHPESPAIKALM
TWAYDLNTNAPLHDAANSJNPAHSGHVDPKALSPGLVDYDISIEEYIRFCSL
DYTVDHIVAIVKPSVNCXKPSDFQLNPSFSLFGKRVVRYTRVTVNVAASSV
YKVTVNGAPSVGISVSKPSKLSFKSVGEKKRYTVTFVSKKGVSMTNKAEFGSITWNPQ
HEVRSPVAFPSWNR"
2288..2541
/gene="At2g05920"
misc_difference 2524..2525
/gene="At2g05920"
/note="not present in genomic sequence"
BASE COUNT 617 a 549 c 600 g 775 t
ORIGIN
Query Match 20.7%; Score 481.8; DB 8; Length 2541;
Best Local Similarity 55.3%; Pred. No. 8.6e-126;
Matches 1033; Conservative 0; Mismatches 812; Indels 24; Gaps 4;
QY 380 ACTCCGGTGTATGGTCTAAATCTCGGTTTGGTCAAGGCACATTTATCGCGTGCITGATA 439
Db 366 AATTCGGTGTTCACGATCTCGGTTCTTCTTAACGGCGTTATCATCGGAGTTTAGATA 425
QY 440 CTGGAGTTTGGCTGAAGTCTAGCTTTGACGATACCGGAATGCTTTCGATTCACGGA 499
Db 426 CTGGCGTTTGGCTGAATCTAGAGCTTCGATGATGATGATGATGATGATGATGATGAT 485
QY 500 AATGAAAGGGATTTGCCAAGAGGAGAAAGTTTCAGTTCITTCGAGCTGTAAACCGGAAGC 559
Db 486 AATGAAAGAGAGATGTGATCTGGTTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 545
QY 560 TAATCGGTCTAGATTTCTATCAGAGGACACCGTGTCTGCTTAATTCACGAGGATCATC 619
Db 546 TTATCGGAGTAGAAGCTTCTCCAAAGGATTTCAAATGGCTTCTGTTGGTGGTGGTGGTGG 605
QY 620 CAACAATGCTGTGTGATACATTTCCGCAAGAGATTCACGGGACACGGGACTCACACCG 679
Db 606 GTAA-----GCGTGAATCTGTTCTCTCGTGTGTTGACGAGATGGAACACATACAT 659
QY 680 CCTCAACAGTTGTTGGATCTCTGTTTCGATCGGAATGTTCTTGGCAATGGAGCTGGTG 739
Db 660 CAACACTACCGCGCGGATCCGCGTTAGAAACGCTAGCTTCTCGTTACGCGCGGCTA 719
QY 740 TGGCTCGTGGGATGGCTCCTCGGAGCTCACATTCGAGTCTATAAAGTCTGTTCGTTCAATG 799
Db 720 CGGCCAGAGTATGGCCACTCGTGTCTGTTGTTGCTACTTATAAGTTTCTGGAGTACTG 779
QY 800 GTTGTGTACAGTCTGATCATTCATGACAGCTATAGATGTAGAGTATCAAGTAAAGTCGATG 859
Db 780 GTTGTGTGATCTGATATCTAGCTGTCTATGATGATGATGATGATGATGATGATGATGATG 839
QY 860 TTCTTTTCGCTTTCCCTTGGCGGTTTCCCTATTCTTGTATGATGATGATGATGATGATGATG 919
Db 840 TGCTTTTCGTTATCTCTTGTGGTGGTTCGTCGGTATTATCGGATACGATTCGATTCGATTC 899
QY 920 GAACATTCGAGCCATGGAAACGCGGTATATCTGTAATCTGTGACGCTGTGTAACAAACGGTC 979

```

900 GAGCGTTTTCGGCTATGAGAGAGGTTGTTTTTGTGTTCTGTTCTGCTGCTTAATAGTGGTC 959
980 CAATCGAAAGCTCTGTTGCAACACACAGCTCTCTGGGTTCTCAACATTTGGCGGAGGACGC 1039
960 CTACTAGAGCTTCTGTTSCCAATGTTCTGCTCTCTGGGTTATGACTGTTGGTCTGGTACTT 1019
1040 TTGATCGAAGATTTCCCGCTGTGTCAGATTAGCCACGGAAGGCTTCTCTATGGAGAGT 1099
1020 TAGATAGAGATTTTCGGCTTTTTCGGAATCTCGGTAACGGGAACGACATACCGGTGTTT 1079
1100 CATTGTATCCGGGAAAGGTATAAAGAAATGCCGGGAGAGAGGTTGAGGTGATTTACGTCA 1159
1080 CGCTGTATAGCGGTGTAGGAATGGGA-----CGAAGCCGCTTGAATTTGGTTTATAATA 1133
1160 CAGAGAGAGATAAAGGAGTCACTTCTGTTTGAGAGGCTCACTTCCAGAGAGAATCC 1219
1134 AAGGGAATAGTAGTTCGAGTAATCTTTGTTTAACTCGCTTTCGCTTGATTCGAGTATTGTTT 1193
1220 GAGCAAAATGGTATTGTTGATCGCGAGTCAATGGAAGATCGGAAAGGAGAGAGCGG 1279
1194 GTGGGAAGATTGTTGTTGATAGAGGTGTAATGCTAGAGTTGAGAAAGGAGCTGTGG 1253
1280 TTAAGAGCTGGAGGAGTTGCAATGATCTTAGCAATACAGAGATCAACCAAGAGAAG 1339
1254 TTAGAGATGCTGGTGGTTAGGATGATAATGGCGAATCTGCTGCGAGTGGAGAGGAGC 1313
1340 ATTCTATTGAGCTTCACTCTTTACAGGTACATTTGATTTGTTACATGAGTCACTGCTTC 1399
1314 TTGTGGCGGATGATCAITTTGCTTCCCGCATCGCTGTAGGGAAGAGACTGGTGAATTAC 1373
1400 TGAAGGCTTATGTTAATGCCACGCTGAAACCAAGGCGCGATAATTTTGGTGGTACGG 1459
1374 TTAGGAGTATGTTAAGTCAGATCTTAACCAACCGCTCTTCTGTTTTTAAAGGAACGG 1433
1460 TCATTGGGAGTCAAGACACCGAGTGGTCACTTTTTCAGCTCGAGGACCGAGTTTAG 1519
1434 TTTCTGAGTTAAGCGCTCTCTGTTGTTGCTGCTTTTAGCTCGAGAGTCTTAATACTG 1493
1520 CCAATCTTCGATCTAAACCGGATATGATGTTCCGGGAGTCAATCATTTTCGGCTT 1579
1494 TTACTCTGAAATCTGAAGCTGATGTTATTGTTCTCGAGTTAATATTTCGGTGGTT 1553
1580 GGCCTCAAAATCTAGGACCAACCGGACTCTCTTATGATTCAGAGAGTAACTTCACTG 1639
1554 GGTCTGACGCTATTGGTCTCTACTGCTTGTGACAGGAGTCTTAGAGAGTCACTCAACA 1613
1640 TAATGTCAGGAACCTCAATGCTTGTCCACATGTTAGCGGAATCACTGCTCTTATCCGCT 1699
1614 TCATGTCAGGTACGTCAATGTCATGCCACACATCAGTGGTTTAGCGGCTTTTGAAG 1673
1700 CTGCATACCCGAACTGTTCTCCAGTCCGAATCAATCCGCTTGTATTCAGAGAGTAACTTCACTG 1759
1674 CAGCTCACCTGAGTGGAGTCCGAGTGTCTATCAATCAGCTCTCATGACTACAGCTTAAG 1733
1760 TGTAGTCTGTCAAGGGAAGAGTAAGGATGTT-----ACAAACCCAGCGGTGTT 1813
1734 TTTCTTGACACACCAACCACTCTCTCCATGATGCTGCAGACACAGAGCTATCTAACCCAT 1793
1814 TTGCGATTGGAGCGGCGATGTAATCCGAAAGGCGATAAACCAGGATTTGTTTACA 1873
1794 ATGCTCAGCGTCCGGGCCATGATGATCCCAAAAGGCTCTCTACAGGCTTTGTTCTAGC 1853
1874 ACATTCACCAAGTGAATTACATACTTCACTCTGCTGCTCTCTGTTGATTCACAGATCAGATA 1933
1854 ACATCTCAACCGAGGAATACATCAGTTTGTGCTCTCTAGACTACACAGTCCATCACA 1913
1934 TTTTAGCAATCACTCATAGAACGTGAGTCAATGGAATATTCGGGAAACCCCGGTT 1993
1914 TTGTTTGGAT-----TGTGAAGCGACCTAGCGTTAACTGCTCGAAGAGTTCTCAGATC 1967
1994 TTAGTCTCAATTACCGGTCGATAGCGGTGATTTTCAACGTTGGCAAGTACGAGATGA 2053
1968 CTGGTCAGCTCACTACCAAGTTTCTCGGTTTTGTTGGGGTAAAGAGTTGTCGGT 2027

QY 2054 TCACAAGCGGTGTCACATAACGTTGGAGTCTTACTCGATATCTCAGTGAATGTCAGG 2113
Db 2028 ACATCGGAAGTAAACAATGTTGTCAGCAACGTCGGTTTACAAAGTGACGGTTAATG 2087
QY 2114 CTCAGAGGGGATCAAAAGTATTGTTCAATCTTAAGAGACTTGTGTTCAACACGTGATC 2173
Db 2088 GAGCTCTAGTGTGGGAATCTCTGTTAAACCATCGAACTTTCGTTTTAAAGCGTGGAG 2147
QY 2174 AGACGCTCAGCTATAGAGTATGTTTCTTATTGAGAGAGAAAACAGAGGAGGGAGGTGG 2233
Db 2148 AGAAGAGAGGTACACAGTCACTGTTGTTAGCAAGAAGGAGTGTGATGCGAACAGG 2207
QY 2234 CTAGCTTTG 2242
Db 2208 CTGAGTTTG 2216

RESULT 10
BT001082
LOCUS Arabidopsis thaliana At5g67360/X8K14_8 mRNA linear PLN 18-OCT-2002

DEFINITION Arabidopsis thaliana (thale cress)
ACCESSION BT001082
VERSION BT001082.1 GI:24111424
KEYWORDS FLI CDNA
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
1 (bases 1 to 2274)

REFERENCE
AUTHORS Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
Chan M.M., Chang C.H., Dale J.M., Havaehizaki Y., Hsuan V.W.,
Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J.,
Lam B., Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M.,
Pal M.C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C., Yamada K.,
Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A. and
Ecker J.R.

TITLE Arabidopsis ORF clones
JOURNAL Arabidopsis
REFERENCE 2 (bases 1 to 2274)
AUTHORS Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
Chan M.M., Chang C.H., Dale J.M., Havaehizaki Y., Hsuan V.W.,
Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J.,
Lam B., Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M.,
Pal M.C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C., Yamada K.,
Yu G., Yuan S., Shinozaki K., Davis R.W., Theologis A. and
Ecker J.R.

TITLE Direct Submission
JOURNAL Submitted (18-OCT-2002) Salk Institute Genomic Analysis Laboratory
REFERENCE (SIGNAT), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA : RIKEN
Arabidopsis Full-length cDNA) : Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Havaehizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and
sequenced the PUNI (ORF) clones using the RAFL cDNAs: Cheuk, R.,
Chen, H., Kim, C.J., Shinn, P., Ban, J., Bowser, L., Chan, M.M.,
Chang, E., Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G.,
Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S.,
Pal, M.C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M.,
Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Davis, R.W., Theologis, A.,
and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
contributed equally to this work as PIs.

FEATURES		Location/Qualifiers
source		1..2274
		/organism="Arabidopsis thaliana"
		/mol_type="mRNA"
		/db_xref="taxon:3702"
		/chromosome="5"
		/clone="U21848"
		/note="This clone is in puni 51"
		ecotype: Columbia
CDS		1..2274
		/note="cucumisin-like serine protease (gb AAC18851.1)"
		/codon_start=1
		/product="At5g67360/k8k14_8"
		/protein_id="AA046863.1"
		/db_xref="GI:24111425"
		/translation="MSSSFLSSSTAFFLLLCGRCHVSSSSDQGTIVHMAKSMQSSS FDLSNNYISDSASLELYTENAIFGSTRLTQEEADSLMTQPGVISVLPBHR YELHTTLPFLDEHDTADLFPFAGSYDVVGVLDLTGVMPESKSGPFGFIPSS WKGCEAGTFLASLCNRKLIAGFFARFYESTMGPIDESKSRSPRDDHGHHTSS TAAGSVGGASLLYAGGTARGMAPRAVAVYKVMGLGCFSDIILAAIDKADNVN VLSNLSGGNSDYVDGVAIGAFAPAMERGIILVSCAGNAGPSSSSLSNVAPWITTVGA GLDRDPALLNGNKNFTGVSLFKGEALPDKLLPFIYAGNASNATGNLMTGTLI PEKVKIKVMDRGINARVQKGVVKAAGVGMILLANTAANGBELVADAHLLPATTG EKAGDIIRHYVTPNPNTASISILGTGVGKPSVFAAFSSRGPNSITPNILKPDLIA PGVNIILAAATCAAGTGLASDRRVEFNIIISGTHGSHVSPPTATNPGLIYDLTETVSL IRSAIMTAYKTYKDGKPLLDIAFGKSTPFDHGHVSPPTATNPGLIYDLTETVSL GFLCALNYTSPQIRSVRRNYTCDFSKSYVADLNYPSFAVNVGVGAYKYTRTVSV GGAGTYSVKVTSSETTGKISVEPAVLNFKANERKSYITVTFVDDSKPSSNSFGSIE WSDGKVVGVSPVA:SWT"
BASE COUNT		535 a 566 c 571 g 602 t
ORIGIN		
Query Match		20.6%; Score 480.4; DB 8; Length 2274;
Best Local Similarity		56.2%; Pred. No. 2.1e-125;
Matches 973; Conservative		0; Mismatches 741; Indels 18; Gaps 3;
QY	210	ACCTCTCTACTCTGCTCTGCGATTGAAGGATTTGCTGCTCAGTTGACTGAATCAGA 269
DB	192	ATTGCTCTACACTTACGAGACGGGATTCATGGATTTCTCAACACGGCTAATCAAGAAGA 251
QY	270	AGCCGAGATCTGAGATATTACCTTGAAAGTTGTTGCACTGAGACCTGACCACTGTTCTCA 329
DB	252	AGCCGACTCGCTCATGACTCAACCTGGTGTATCTCGTTTACGGGACACCGTTACGA 311
QY	330	GATTCAAAACCACTTACTCTTACAGTCTTGGGACTCGACCGGTTTGGAACTCCGGTGT 389
DB	312	GCTACACACCACTCGTACTCTCTCTCTCGGCTCGACG---AACACACCGCAGATCT 368
QY	390	ATGGTCTAAATCTCGTTTGGTCAAGGCACATATTCGGCGTCTTGATCTACTGGAGTTG 449
DB	369	GTTCCTGAAGCCGCTCTTACAGCGACGTGCTGTCGGAGTTCTCGATACCGGAGTTG 428
QY	450	GCTGAAAGTCTAGCTTTGACGATACCGGATGCTTCGATTCCACGGAAATGAAAGG 509
DB	429	GCCGAGAGTAAAGCTACTCCGACGAGGATTCGGTCTTATCTCTCTCTCGTGAAGG 488
QY	510	GATTTGCCAAGAGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGCTAATCGGTGC 569
DB	489	CGGATCGGAGGCGGAAACCAATTTACCGCTTCTCTGTAAACCGTAAACTAATCGGAGC 548
QY	570	TAGATTCTTCATCAGAGACACCGTGTGCTAAATTCACAGAGGATCACCAAAACATGCC 629
DB	549	AAGATTCTTCGCTCGTGGTTAC-----GAATCAACGATGGGACCAATCGATGAATC 599
QY	630	TCGTGAATACATTTCCGGAAGAGATTCAACGGGACACGGGACTCACACCGCTCAACAGT 689
DB	600	TAAAGAAATCAAGATCTCTTAGAGACGACGACGGTCACCGAACTCACCTCATGACCGC 659
QY	690	TGGTGATCTCTGTTTCGATCGCGAATGTTCTTGGCAATGAGCTGGTGTGGTCTGTGG 749
DB	660	CGCTGGATCCGTGTTGAAGGAGCTAGCTTATTAGGTACGCTTCAGGAAACAGCTCGTGG 719
QY	750	GATGGCTCTCGGAGCTCACATTCGAGCTATAAAGTCTGTGGTTCAATGGTTGTTCAG 809
DB	720	TATGCTCCACCGCGCTGCTGCTGCTGCTTAAAGTCTGTTGGCTCGGTGTTGTTTCAG 779
QY	810	CTCTCACATTTCTAGCAGCTATAGATGTAGCCATTCAAGATAAAGTCGATGTTCTTCGCT 869
DB	780	CTCACATATTTTAGTGTCAATCGATAAAGCCATCGCGATACGTCATGCTCTGCTCGAT 839
QY	870	TTCCCTTGGCGGTTTCCCTATTCCTTTGTATGATGACACAATCGGCATGTGAACAATCCG 929
DB	840	GTCACTCGGCGTGGAAATGTGCGGATTAATATAGACACGGTGTTCGATCGGAGCATTCGC 899
QY	930	ACCCATGAAACCGCGGTATATCTGTAATCTGTGCAGCTGGTAACAACGCTCCAATCGAAAG 989
DB	900	CGCCATGGAAGAGGAGGATTTTAGTATCTTGCTGAGCTGGTAATGCTGTGCGAGTAGTTC 959
QY	990	CTCTGTTGCAACACACAGCTCCTTGCTGCTCAACCAATTCGCGCAGCAGCAGCTTGATCGAAG 1049
DB	960	TAGTTTATCAACAGTAGCTCCATGATCACAACCTCTGCTGCTGCTGCTTATGATCGTGA 1019
QY	1050	ATTTCCTCGCTGTGCTCAGATTAGCCACCGAAAGCTTCTCTATGAGAGTCAATGTATGCC 1109
DB	1020	TTTTCCGGCGCTTTCGATTTCTCGGTACCGCAAGAAATTTCAACCGAGTTCTTCTGTTTAA 1079
QY	1110	GGGAAAAGGTATAAAGATGCGCGGAGAGAGAGTTGAGGTGATTTACGTCAACAGGAGAGA 1169
DB	1080	AGGAGAAGCTTTTACCGGATAAATTTCTGCGCTTTATTTACGCTGGGAATGCTAGTAATGC 1139
QY	1170	TAAAGGAGTGTGTTCTGTTTACGAGGCTCACTTCCAGAGAGAAATTCGAGGCAAAAT 1229
DB	1140	TACTAATGTTAATCTCTGTATGACCGGAACTTTGATCCCGGAGAAAGTAAGGGGAAGAT 1199
QY	1230	GGTATTTCTGATCGCGGAGTCAATGGAAGATCGGAGAAAGAGAGAGCGGTTTAAAGAAAGC 1289
DB	1200	TGTGATGCTGATAGAGGAATTAATGCTAGAGTTTCAGAAAGGTGATGTGTTTAAAGCAGC 1259
QY	1290	TGAGAGGTTGCAATGATCTTAGCCAAATACAGAGATCAACCAAGAGAGAGATCTTATGTA 1349
DB	1260	AGGTGAGTTGGAATGATTTCTGCTTAACTCTCGGCGAATGTTGAGAGCTTTGTTGCGGA 1319
QY	1350	CGTTCTATCTTTACCACTGATGTTGTTGTTTACACTGAGTCAGTCCCTTCTGAAGCTTTA 1409
DB	1320	TGCTCATTTTGTACCGCGACACCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1379
QY	1410	TGTTAATGCCACGGTGAAACCAAGCGCGGATTAATTTTGGTGTGTTGTTGTTGTTGTTGTTG 1469
DB	1380	CGTGACTACTGATCTTAATCTCCACCGCTTCGATTTCAATCTTTAGGAAACAGTCTGCTGCT 1439
QY	1470	GTCAACGAGCACCGGAGTGGCTCAGTTTTCAGCTCGAGSACCGAGTTTAGCCAAATCCCTTC 1529
DB	1440	TAAACCATCTCCGGTTGTCGACGCTTTAGCTCACGTGACCGAATTCGATTACACCGAA 1499
QY	1530	GATCTAAACCGGATATGATTGTTCCGGGAGTCAATATCATTTGCGGCTTGGCTTCAAAA 1589
DB	1500	TATTTTAAACCGGATCTGATCGCTCTGAGTAAACATCTCTCGCGGCTGAGCCGCTGC 1559
QY	1590	TCTAGAACCAACCGGACTTCTTATGATTCAGAGAGTTAACTTCACTGTAATGTCAGG 1649
DB	1560	TGCTGACCAACCGGACTCGCTTCGATTTCTCGCGCTGGAGTTCAATATCATCTCTGCG 1619
QY	1650	AACTTCAATGCTCTTGTCACATGTTAGCGGAATCACTGCTCTTATCCGGTCTGTCATACC 1709
DB	1620	GACCTGATGTTCTTGCTTCACTGTTAGTTGTTTACGGGCTTCTCAAGTCGCTGATACC 1679
QY	1710	GAACTGCTTCCAGCTGCAATCAAAATCCGATATGATGACAAACAGCGGATTTGATGATCG 1769
DB	1680	TGAATGTAGCCCGCGCGATTAGATCGCGCTTATGACCAACCGCTTACAAAACCTTACAA 1739
QY	1770	TCAGGGAAGCGATA-----AAGATGTTAACAACACCGCGGCTGTTTTCGCTGCG 1823
DB	1740	AGACGGTAAACCGCTTACTCGACATCGCAGAGGAAGCCCTTCAGCCCGTTCGATCACCG 1799
QY	1824	AGCAGGCAATGTGAATCCCGCAAAAGCGATTAACCGGATTTGTTTACCAATTCATCAAC 1883

QY 630 TCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCCCTCAACAGT 689
Db 631 TAAAGAACTCAAGATCTCCTTAGAGACGACGAGCGTTCACGGAACTCACACCTCATCGACCGC 690
QY 690 TGGTGGATCCTCTCTTTCCGATGGGAATGTTCTTGGCAATGGAGCTGGTGTGGCTGGTG 749
Db 691 CGCTGGATCCGCTGTTGAGGAGCTAGCTTATTAGGCTAGCTTACAGGAACAGCTGGTG 750
QY 750 GATGCTCTCTGGAGCTCACATTCGAGTCTATAAAGTCTGTGGTTCAATGGTTGTTACAG 809
Db 751 TATGGCTCCAGCGCTCGTGTGCTGTTTACAAAGTCTGTGGCTGGTGGTGTGTTTCAG 810
QY 810 CTCTGACATTCAGAGCTATAGATGAGGATTCAGGAATTCAGATGAAGTGCATGTTCTTCGCT 869
Db 811 CTCAGATATTTAGCTGCAATCGATAAAGCCATCGCGGATAAGCTCAATGCTCTGTCGAT 870
QY 870 TTCCTTTGGCGTTTCCCTATTCCTTTGTATGATGACACAATCGCCATTCGACATTCG 929
Db 871 GTCACTCGCGGTGGAATGTCGGAATATTATAGAGACGGTGTTCGATCGGAGCATTCGC 930
QY 930 AGCCATGGAACGCGGTATCTGTAATCTGTGACGCTGGTAACAACCGTCCAATCGAAAG 989
Db 931 CGCATGGAAGAGGATTTAGTATCTTGTCTGACGCTGGTAATGCTGTCCGAGTAGTTC 990
QY 990 CTCTGTTGCAACACAGCTCCTCGGTCTCAACCATTCGCGCAGGACGCTTGAATCGAAG 1049
Db 991 TAGTTTATCAACGCTAGCTCATGATGATCAACACTGTTGCTGTGTTAGTACTTATAGATCGTA 1050
QY 1050 ATTTCCCGCTGTGTGAGATAGCCACGGAAGCTTCTCTATGGAGAGTCAATGATCC 1109
Db 1051 TTTTCCCGCGCTTCGATTCCTCGTAAACGGAAGATTTCAACCGAGTTCCTTTGTTTAA 1110
QY 1110 GGGAAAGGTATAAGATCCGGGAGAGAGTTGAGTGATTTACCTCAGCAGGAGGA 1169
Db 1111 AGGAGAGCTTTACCGGATAAATTTGTCGCTTTATTACGCTGGGAATGCTAGTATGC 1170
QY 1170 TAAAGGAGTGAATCTCTTTGAGAGGTCACITCCAGAGAGAAATCCGAGCGCAAAAT 1229
Db 1171 TACTAATGTAATCTCTGTATGACCGNACTTTGATCCGAGAGAAATGAAGGGAGAT 1230
QY 1230 GGTGATTTGTGATCCGCGAGTCAATGGAAGATCCGAGAAAGGAGAACGGTTAAAGAAC 1289
Db 1231 TGTGATGTGTATAGAGAAATTAATGCTAGAGTTTCAAGAGGTCATGTGTTAAAGCAGC 1290
QY 1290 TGGAGGATTCGATGATCTTAGCCATACAGATCAACCAAGAAAGATCTATTGA 1349
Db 1291 AGGTGGAGTTGGAATGATTTCTGGTAAACACTCGCGCAATGGTGAAGAGCTTGTTCGGA 1350
QY 1350 CGTTCATCTTACAGCTACATGATTTGTTTACACTGAGTCAGTCTCTTCTGAAGGCTTA 1409
Db 1351 TGCTCATTTGTTACCGGCGAACACCGTTGGTGAAGAACCGGCTGATATATCCGGCATTA 1410
QY 1410 TGTAAATCCACGGTGAACCAAGGCGCGGATAATTTTGGTGGTACGGTGAATGGGAG 1469
Db 1411 CGTGACTACTGATCCTAATCCACCGCTTCGATTTCAATCTTAGAACACGTCGTCGCTG 1470
QY 1470 GTCAGAGACCGGAGGTGGTCTAGTTTTCAGTTCGAGGACCGAGTTTACCAATCCTTC 1529
Db 1471 TAAACCATCTCGGTGTCGAGCGGTTTAGCTCAGTGGACCGGAATTCGATTACACCGAA 1530
QY 1530 GATATAAACCGGATATGATGTCGCGGAGTCAATATCATTCCGGCTTGGCTCAAAA 1589
Db 1531 TATTTCTTAACCGGATCTGATGCTCTCGGAGTAAACATCTCCCGCGTGGACCGGTGC 1590
QY 1590 TCTAGACCAACCGGACTTCCTTATGATTCAAGAGAGTTAACTTCACTGTAATGTCCAGG 1649
Db 1591 TGCTGGACCAACCGGACTCGCTTCGATTTCTCGCGGCTGGAGTTCAATATCACTCTGG 1650
QY 1650 AACTTCAATGTCTTGTCCACATGTTTACGCGAATCACTCTCTTATCCGGTCTGCATACC 1709
Db 1651 GACGTGATGCTTCTGCTCCCTCAGTTAGTGGTTTACGCGGCTTCTCAAGTGGTGCATCC 1710
QY 1710 GAACCTGGTCTCAGCTGCAATCCGATTTGATGATCAACACAGCGGATTTGTACGATCG 1769

Db 1711 TGAATGTAGCCGCGGGCGATTAGATCGCGCTTATGACACCGCTTACAAAACCTACAA 1770
QY 1770 TCAACGGGAAAGCATATA-----AAGGATGTAAACAAACCCCGGTGTGTTTGGATTGG 1823
Db 1771 AGACGCTAAACCGTTTACTCGACATCGCACAGGGAAGCTTCGACCGCGTTTCGATCACGG 1830
QY 1824 ACACGGGATGTGAATCCGCAAAAGGCGATAAACCCGGGATGTTTACACATTCACACC 1883
Db 1831 TCGAGGACACGCTGTCCACCAACATGCGCACTAATCCAGGACTCATCTACGATCAACGAC 1890
QY 1884 AGTGATTTACATAAATTAACCTTACCTCTGCACTCTTGGATTACAAAGATCAGATATT 1935
Db 1891 GGAGGATTACTTAGGCTTCCTCTCGGATTGAATTACACATCGCGCGAAATT 1942

RESULT 12

ATARA12 2423 bp mRNA linear PLN 14-MAR-1996
LOCUS A.thaliana mRNA for subtilisin-like protein.
DEFINITION X85974
ACCESSION X85974.1
VERSION GI:757533
KEYWORDS arai2 gene; subtilisin-like protease.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1
AUTHORS Ribeiro, A., Akkermans, A. D., van Kammen, A., Bisseling, T. and Pawlowski, K.
TITLE A module-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development
JOURNAL Plant Cell 7 (6), 785-794 (1995)
MEDLINE 95375542
PUBMED 7647567
REFERENCE 2 (bases 1 to 2423)
AUTHORS Ribeiro, A.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-1995) A. Ribeiro, Dept. of Molecular Biology, Agricultural University Wageningen, Dreijenlaan 3, NL- 6703 HA Wageningen, NETHERLANDS
COMMENT Related sequence: T04180.
FEATURES
Location/Qualifiers
1..2423
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/culturivar="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="lambda-PRL2 (T Newman)"
/clone_lib="cDNA library derived from 4 pools of mRNA, (1) 7 days germinated, etiolated plants; (2) tissue culture grown roots; (3) staged plants half with 24 hours light cycle, half on 16 hours light, 8 hour dark rosettes; (4) as (3) but aerial tissue (stems, flowers and siliques)"
1..2423
/gene="arai2"
<1..2242
/gene="arai2"
/codon_start=2
/product="subtilisin-like protease"
/protein_id="CAA5963.1"
/db_xref="GI:757534"

source

gene

CDS

Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kaniya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

TITLE JOURNAL REFERENCE AUTHORS

Arabidopsis Full Length cDNA Clones
Unpublished
2 (bases 1 to 2549)

Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kaniya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Direct Submission
Submitted (19-FEB-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kaniya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

FEATURES source

Location/Qualifiers
1..2549
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFL07-08-N18 (R10656)"
/note="this clone is in a modified pBluescript vector (pLC-1) as a BamHI/XhoI insert."
ecotype: Columbia

gene

1..2549
/gene="At5g51750"
1..59
/gene="At5g51750"
60..2402
/gene="At5g51750"
/codon_start=1

5' UTR

/evidence=experimental
/product="putative subtilisin serine protease"
/protein_id="RA187307.1"
/db_xref="GI:19424032"

CDS

/translation="MANKNPLQK9PLFILLINLIPLQAEITTTQISTKTKVTHMDKS
AMPLPYTHVLQWYSKINSVTCXKQSEEGNNRILTYQAFHGLAQLTQEAERL
EEDVDGVAIVETRYELHTKSPFLGLERQESRWAERVTDHVVVDLTGTWPE
SEFNDTGMSPVATWACETGRFLKNCNRKIVGARVFRGYEATKQKIDDEEY
KSPDRDGHGTHATVAGSPVKGANLFGPAYTARGMAQKARVAAYKVCWVGCSFS
DILSAVDQAVAGVQVLSISLGGVSTVSRDLSIATFGAMEMGVFVSCSAGNGPDP
ISLTVNSWITTCGLDNRDRFPATKIGTWRTKGVSLYKGRVLPKNQYPLVILG
RNASPDFTSCGLDNRDRFPATKIGTWRTKGVSLYKGRVLPKNQYPLVILG
NGEELVADSHMLPVAVGEKEGKIKQYAMTSKATASLEILGTRIGIKPSVVAAS
SRGNFLSLEILKPDLLAPGVNLIAMTGDMAPSLSLSDPRVFNILISGTSVVAAS
SGVAALIKSRHPDMSPAIKSALMTAVYDHNMFKPLTDAAGAAPSPFYDHGAGHIDP
LRATDPLGVLDIGQVEYEFPLCTQDLSPSQLKVFTHKSNRTCKHTLAKNPGLNYPAL

SALFPFTHVKAMTLRTVTVNVPHTSIYSKYVSPFPKGSVTVQPKTLNFTSKHQKLS
YTVFRTFRMKRPFEGVLVWKSITTHKVRSPVITWLPPL"
2403..2549
/gene="At5g51750"
BASE COUNT 765 a 521 c 564 g 699 t
ORIGIN

Query Match 19.2%; Score 447.4; DB 8; Length 2549;
Best Local Similarity 52.9%; Pred. No. 5.7e-116; Indels 30; Gaps 4;
Matches 1073; Conservative 0; Mismatches 926;
185 AAGAAGAGGAAGAGCGCTTCTTCGACATCTCTACTCTCTATGGCTCTCGCATGAAGGAT 244
268 AAGAAGAGGAAGGTAATAACAACAGGATCTACATACAGACTGCTTCCACGGTT 327
245 TTGCTGCTCAGTTGACTGAATCAGAAGCGGAGATCTGAGATATTCACCTGAAGTTGTTG 304
328 TAGCAGCTCAGCTTACTCAAGAAGAAGCAGAGAGGCTTGAGGAAGAAGATGCTGTGTAG 387
305 CAGTGAGACCTGACCATGTTCTTCAGGTTCAACACCATCTCTTACAGTTCTTGGGAC 364
388 CTGTGATACCTGAGCAAGATACGAGCTTACACACAGAGTCCACAGTCCACGTTCTTGGGT 447
365 TCGACGGTTTTCGAAAACCTCCGGTGTATGCTTAAATCTCGGTTTGGTCAAGGCACAAATTA 424
448 TAGAAGACAAGAAGTGAAGAGTTGGGCTGAGAGATCACCAACCATGATGTGGTAG 507
425 TGGGCTGCTTGTATCTGAGTGTGGCTTGGCTTGAAGTCTTACGATACCGAATGC 484
508 TTGGTGTATAGACACTGTAICTGGCTGAGAGTGAAGCTTCAACGATACAGTATGT 567
485 CTTCGATTCACCGAAATGGAAGGATTTGCCAAGAGGAGAAAGTTTCAGTTCTTGA 544
568 CCCTTCTCTCTACTTGGAGAGGAGCTTGTGAAACTGGAAAAAGATTTCTTGAACGTA 627
545 GCTGTAAACCGAGCTTAATCGGCTGATGTTCTTCAATCAGAGACACCGTGTGCGTAAT 604
628 ACTGCAATAGAAGATCGTGTGTAGAGTTTCTATAGAGG-----CTATGAAG 678
605 CACCAGAGGAATCACCAACATGCTCGTGAATACATTTCCGCAAGAGATTCACAGGGAC 664
679 CTGCAACGGGAAGATCGATGAAGAGCTTGATATAAGTCAACGAGAGACAGATGGTC 738
665 ACGGAGCTCACACCGCTCAACAGTTGGTGGATCTCTGTTTTCGATGGCAATGTTCTTG 724
739 ACGGAGACACACTGCAGCTACTGTAGCTGGCTCACCTGTTAAAGGAGCTAATCTTTTG 798
725 GCATGAGAGCTGTGGCTCGTGGATGGCTCTCGGAGCTCACATTGCACTATATAAG 784
799 GATTGCTTATGGACAGCTCGAGGATGGCTCAAAAGGCTAGAGTTGCTGCTTATAAG 858
785 TCTGTTGGTTCATGGTTCTTACAGCTCTGACATCTACAGCTATAGATGTAGCGATTC 844
859 TCTGTTGGTTCGAGGGTGTTCAGTTTTCAGTTCAGACATTTGTCGGCTGTGTATCAAGCTGTTG 918
845 AAGATAAAGTCGATGTTCTTTCGCTTTCCGCTTGGGGTTCCTTATTCCTTTGTATGATG 904
919 CTGATGAGGTTCAAGTACTCTCTATATCATTAGTGGTGGGCTCTCTACTTATCTTAGAG 978
905 ACACAAATCCCAATTTGGAACATTTCCGAGCCATCGAAGCGGATATCTCTGTAATCTGTGAG 964
979 ACAGTTTGTCTATAGCAACATTTGGAGCAATGGAGATGGGAGTTTCGTTTCTGTTCTG 1038
965 CTGGTAACACCGTCCAAATCGAAGCTCTGTTGCAACACAGCTCCTTGGGCTCAACCA 1024
1039 CCGGTAATGAGGCTCTGATCCGATTAGTCTCACTAATGTTTTCATGATGATCAACAG 1098
1025 TTGCGCCAGGCAAGCTTGTGATCGAAGATTTCCGCTGTGGTGTAGATTAGCAACGGAAGC 1084
1099 TTGCTCAAGTACTATGGATAGAGATTTCCAGCAACAGTGAAGATAGGAACATGAGAA 1158
1085 TTCTCTATGAGAGTCAATGTATCCGGGAAAAAGGTATAAGAAATGCCGGGAGAGAGTTG 1144

Db 1159 CATTCAAAGAGTGTCATCTTACAAAGGAGAGACAGTTTTCCTTAAGATAAACAAGTATC 1218
Qy 1145 AGGTGATTTACGT-----CACAGGAGGAGATAAAGAAAGTGTCTGTTTGAGAG 1195
Db 1219 CTCTGGTTTACTTAGGAAGGATGCAAGTAGTCTGTGATCCAACTCGTTCTGTCTAGATG 1278
Qy 1196 GGTCACTTCCAAAGAGAAGAAATCGAGGCAAAATGGTGAATTTGTATCGCGGAGTCAATG 1255
Db 1279 GAGCTTTGGATCGCGGCATGTAGCGGAAAGATCGTGATCGACCGCGGTGTACTC 1338
Qy 1256 GAAGATCGGAGAAAGGAGAGCGGTTAAAGAGCTGGAGGAGTGAATGATCTTAGCCA 1315
Db 1339 CACGTGTCAAAGGGTCAGGTTGTGAAAGAGCTGTGGAATGGGATGGTTTTAACTA 1398
Qy 1316 ATACAGAGATCAACAAGAGAAATCTATTGAGCTTCATCTCTACCAAGTACATTTGA 1375
Db 1399 ACATCGAAACAATGGTGAAGAGCTTGTTCGAGATTTCTCATATGCTTCCAGCTTTCGAG 1458
Qy 1376 TTGTTTACACTGAGTCAGTCTCTCTGAAGGCTTATGTTAATGACCGGTGAACACCAAGG 1435
Db 1459 TTGAGAGAGAAAGGTAACATAATCAACAGTACGGGATGACGAGTAAAGGAGGACAG 1518
Qy 1436 CGCGGATATTTTGGTGTGACGTTGATGGAGGTCACGAGCACCGGAGTGGCTCAGT 1495
Db 1519 CGAGTTTAGAGATTTCTTGGAAACAAGAAATGGTATCAAACTTCCACAGTTGTAGACGCT 1578
Qy 1496 TTTCAAGTCGAGGACCGAGTTTACCAATCTTCGATACTAAACCGGATATGATGCTC 1555
Db 1579 TCTCTTCAAGAGGACCAAAATTTCTGCTTTGGAGATCTTGAACACGAGATTGTGGCTC 1638
Qy 1556 CGGAGTCAATATCATTTGCGCTTGGCTTCAAAATCTAGGACCAACCGGACTTCCCTATG 1615
Db 1639 CAGGAGTGAATATTTTGCACCTGGACTGGAGACATGGCACCATCGAGTTTATCATCTG 1698
Qy 1616 ATCAAGAGAGTTAACTTCTACTGTATGTCAGGAATCAATGCTCTTGTCCACATGTTA 1675
Db 1699 ATCAAGAGGAGTTAAAGTTCAATATCTGTCTGAACTTCAATGCTATGCTCTCATGTA 1758
Qy 1676 CGGGAATCACTGCTTATCGGCTGCTGATACCGGACTGCTCCAGCTGCAATCAAT 1735
Db 1759 GTGGTGTAGCTGCTTTGATCAAAATCAAGGCATCCAGATTGGAGTCTTCGACGATCAAT 1818
Qy 1736 CGCATTCATGACACACGAGATTTGACGATCTGCAAGGAAAGGATTAAGATGGTA 1795
Db 1819 CAGCTCTATGACAACTGCTTATGTTCTATGACACATGTTTAACTCTTACGATGAT 1878
Qy 1796 ACAAACCGACCGT-----GTGTTGCGATTGGAGCAGGCGATGTGAATCCGCAAAAG 1849
Db 1879 CAGGAGCAGCTCTCTATCGGCTTATGATCAGGTCGAGGACATATAGATCCTTTAAGAG 1938
Qy 1850 CGATTAACCGGAGTTGTTTACACATTCACAGTGGATTACATTAACCTTACTCTGCA 1909
Db 1939 CTACAGATCCTGGTTGGTCTACGACATTTGACCTCAAGATATTTTGAATTCCTCTGCA 1998
Qy 1910 CTCTTGGATTACAGAGATCAGATATTTTAGCAATCACTCATAGAACGTCGAGTGCATG 1969
Db 1999 CTCAGATTTAGTCCATCAGCTTAGGTTATCAAAAACATTCAAAACAGACCTGCA 2058
Qy 1970 GAATATTGCGGAAACCGCGGTTTTAGTCTCAATTTACCTCCGTCGATAGCCGTTGTTTCA 2029
Db 2059 AACACACTTTGCCAAGAAATCCGGAAACTTGAACCTACCGCGGATATCAGCTTTGTTCC 2118
Qy 2030 AACGTGGCA-----AGACTGCGAGATGATCAAGCGTGTCACTAACTGTTGGAGTC 2083
Db 2119 CAGAGAACACATGTTTAAAGCTATGACACTTAGAAGAACAGTGAACCAATGTTGGTCTC 2178
Qy 2084 CTAATCTGATATCTACTGATGATGTCAGGCTCCAGGGGATCAAAAGTTATTGTCATC 2143
Db 2179 ACATTTCAAGTACAAAGTTTCTGCTCGCAATTCAAAGGCGCATCCGTTACTGTCAGC 2238
Qy 2144 CTAAGAGACTTGTGTTCAACACACGTTGATCAGACGCTGAGGTATAGAT 2192
Db 2239 CCAAAACACTCACTTCACTTCGAGACCGAGAGCTTTCCTACAGCT 2287

RESULT 14
AB010074
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE

REFERENCE
AUTHORS

TITLE

COMMENT

FEATURES
source

CDS

AB010074 86212 bp DNA linear PLN 27-DEC-2000
Arabidopsis thaliana genomic DNA, chromosome 5, Pl clone:MIO24.

AB010074 BA000015
AB010074.2 GI:10177866

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsels.

1 (sites)
Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asamizu,E., Miyajima,N.
and Tabata,S.

Structural analysis of Arabidopsis thaliana chromosome 5. IV.

Sequence features of the regions of 1,456,315 bp covered by

nineteen physically assigned Pl and TAC clones

DNA Res. 5 (1), 41-54 (1998)

98290546

9628582

2 (bases 1 to 86212)

Nakamura,Y.

Direct Submission

Submitted (07-JAN-1998) Yasukazu Nakamura, Kazusa DNA Research

Institute, Department of Plant Gene Research; 1532-3, Yana,

Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamura@kazusa.or.jp,

Tel:81-438-52-3935, Fax:81-438-52-3934)

On Sep 15, 2000 this sequence version replaced gi:2760170.

Address for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see

http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?c=MIO24

Genes with similarity to proteins in the databases are described in

'product' or 'note' qualifiers. Genes that have no significant

protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Grail

(Informatics Group, Oak Ridge National Laboratory,

http://compbio.ornl.gov/Grail-1.3/).

GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of

Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and

SplicePredictor (Volker Brendel, Stanford University,

http://grmbln1.zool.iastate.edu/cgi-bin/sp.cgi).

Genes encoding tRNAs are predicted by tRNAscan-SE

(Sean Eddy, Washington University School of Medicine, St. Louis,

http://genome.wustl.edu/eddy/tRNAscan-SE/).

This sequence may not be the entire insert of this clone. It may be

shorter because we remove overlaps between neighboring submissions.

The 5' clone is K10P11 and the 3' clone is MJM18.

Location/Qualifiers

1..86212

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia"

/db_xref="taxon:3702"

/chromosome="5"

/clone="MIO24"

/clone_lib="Mitsui Pl"

1494..2471

/note="gene id:MIO24.19

unknown protein"

/codon_start=1

/evidence=not experimental

/protein_id="BAB11237.1"

/db_xref="GI:10177867"

/translation="MSDMEPKETKPPRKQRPSPVFIWEERPGFKKNQWPSLATF

VESPPLPPPTPVKLVTSVPPFWETPGKELPASNDPQLPPELETATETPLPP

PVPVVKQTSVPFDWEETPGQPCFVDFTSPELLDQPLFPFMKGVEVTSDFDD

ASSDFSVPSSLATNRVTSIGAVDFDNLNTVTSMTPTPAYSDSDTSSTMT

GASLVGASFLKLPRLLPSEKVAASVDQVSTHPLHEVKLTITDMSIGFPV

RTPTQLGELIMMSRRRSYMRRAVENMKQNPVTVRLSNIFSHSLPKPYAP"

Db 24497 CTGTGATACCTGAGACAAGATACGAGCTTCACACTACAAGAAAGTCCAAAGTTTCTTTGGGT 24556
Qy 365 TCGACGGTTTTGGAACCTCGGTGATGTCFAAATCTCGGTTTGTGTCAGGACAAATTA 424
Db 24557 TAGAAAGACAAGAAAGTGAGAGAGTTTGGGCTCAGAGAGTCCCGACCATGATGTGGTAG 24616
Qy 425 TCGGCGTCTTGATACCTGGAAGTTTGGCTGAAAGTCTCTAGCTTTGAGATACCGGAATGC 484
Db 24617 TTGGTGTTTAGACATGGTATCTGGCTGAGAGTGGAGCTTCAACGATACAGGTATGT 24676
Qy 485 CTTCGATTCACCGGAAATGGAAGGATTTGCCAAGAGAGAAAGTTTCAGTTCTTCGA 544
Db 24677 CCCCTGTTCTGCTACTTGGAGAGGAGCTTGTGAAGTCTGAAAGATTTCTTGAACGTA 24736
Qy 545 GCTGTAAACCGAAGCTAATCGGTGCTAGATCTTTCATCAGAGACACCGGTGTCGTAAT 604
Db 24737 ACTGCAATGAAGAGATCGTTGGTCTAGAGTTTCTTATAGAGG-----CTATGAAG 24787
Qy 605 CACGAGAGGAATCACCAAAATCGTCTCGTAATACATTTCCGCAAGAGATTCACCGGAC 664
Db 24788 CTGCAACGGGAAGATCGATGAAGAGCTTGATATTAAGTCACGAGAGACAGATGGTC 24847
Qy 665 ACGGACTCACACCGCTCAACAGTTGGTGGATCTCTGTTTCGATGGGAAATGTTCTTG 724
Db 24848 ACGGACACACACTGCAGCTACTGTAGCTGGCTCACTGTAAAGGAGCTAATCTTTTG 24907
Qy 725 GCAATGAGAGCTGGTGGCTGCTGGGATGGCTCTGGAGCTCACATGTGAGTCTATAAG 784
Db 24908 GATTTGCTTATGGGACAGCTCGAGGAGTGGCTCAAAAGGCTAGAGTTGCTGTATATAAG 24967
Qy 785 TCTGTTGGTCAATGGTTGTACAGCTCTCACATTTAGCAGCTATAGATGTAGCGATTC 844
Db 24968 TCTGTTGGGTCGAGGAGTTTTCAGTTTCAGACATTTTTCGGCTGTGTGATCAAGCTGTTG 25027
Qy 845 AAGATAAGTCGATGTTCTTTTCCGTTTCCGCTGGCGTTTCCCTATTCTTTGTATGATG 904
Db 25028 CTGATGAGTTCAAGTACTCTATATCATTAAGTGGTGGGCTCTACTTATCTTAGAG 25087
Qy 905 ACACAAATGGCAATGGAACATTCGAGCCATGGAACCGGATATCTGTAAATCTGTGAG 964
Db 25088 ACAGTTGTCTATAGCAACATTTGGAGCAATGGAGATGGGAGTTTTCGTTCTGTTCTG 25147
Qy 965 CTGTAACACGGTCCATCGAAGCTCTGTTCGAACACAGCTCTCTGGGCTCTCAACCA 1024
Db 25148 CCGTAATGAGAGTCTGATCCGATTAAGTCTCACTAATGTTTCTCCATGGATCACACAG 25207
Qy 1025 TTGGCGAGGACACCTTGATCGAAGATTTCCGCTGTGGTCAGATTAAGCAACCGGAAGC 1084
Db 25208 TTGTTGCAAGTACTATGATAGAGATTTTCCAGCAACAGTGAAGTAGAATATAGAA 25267
Qy 1085 TTCTCTATGAGAGTCAATGTTATCCGGGAAAGGATATAAGAAATGCCGGGAGAGGTTG 1144
Db 25268 CATTCAAAGGAGTGTCACTTTTACAAAGGCAAGACAGTTTTCGCTAAGAAATAAACAGTATC 25327
Qy 1145 AGGTGATTTAGT-----CACAGGAGAGATAAGGAGTGTCTGTTGTAGAG 1195
Db 25328 CTCTGTTTACTTAGAGAGGAATCAAGTAGTCTCTGATCCAACTCTTCTGTCTAGATG 25387
Qy 1196 GGTCACTTCCAGAGAGAAATCCGAGGCAAAATGGTGAATTTGATGTCGGGAGTCAATG 1255
Db 25388 GAGCTTTGGATCGGGCCATGTAGCGGGAAGATCTGATATGCGACCGGGTGTACTC 25447
Qy 1256 GAAGATCCGAGAAAGGAGAGCGGTTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCCA 1315
Db 25448 CACGTGTCAAAAGGTCAGTTTGTGAAGAGCTGTGGAAATGGGATGTTTAACTA 25507
Qy 1316 ATACAGAGATCAACCAAGAGAAAGATCTTATGAGCTTCACTCTTACCAGCTACATGA 1375
Db 25508 ACCTGCAACAAATGGTGAAGAGCTTGTGAGATTTCTCATATGCTTCCAGCTGTTGAG 25567
Qy 1376 TTGGTTACACTGAGTCACTCTCTGAGGCTTATGTTAATGCCAGCTGAACCAAGG 1435
Db 25568 TTGGAGAGAAAGAGGTAACCTAATCAACAGTAGCGGATGACGAGTAAAGGCGAG 25627

Qy 1436 CGCGGATAAATTTTGTGGTACGGTGAATGGGAGGTACAGACACCGGAGGTGGTCAGT 1495
Db 25628 CGAGTTTAGAGATTTCTTGGAAACAAGAAATGGTATCAAACTTCCACAGATTTGTAGACGGT 25687
Qy 1496 TTTGAGTTCAGAGACCGAGTTTAGCCATCTTCGTACTATAAACCGGATATGTTGCTC 1555
Db 25688 TCTCTTCAAGAGGACCAAAATTTTCTGCTTTGGAGATCTTTGAACACAGACTTGTGGCTC 25747
Qy 1556 OGGGAGTCAATATCATTTGGGCTTGGCTCAAAATCTTAGGACCAACCGGACTTCCCTTATG 1615
Db 25748 CAGGAGTGAATATTTCTTGACGTTGGACTGGAGACATGGGACCATCGAGTTTATCATCTG 25807
Qy 1616 ATTCAAGAAGAGTTAACTTCACTGTATGTGAGAACTTCAATGTCTTGTCCACATGTTA 1675
Db 25808 ATCCAAGAGGGTTAAGTCAATATCTGCTGGAACCTTCAATGTCTATGCTTCTCATGTA 25867
Qy 1676 GCGAAATCACTGCTCTTATCCGGTCTGCATACCGGAATCGGTCTCCAGCTGCAATCAAT 1735
Db 25868 GTGGTGTAGCTGCTTTGATCAAAATCAAGGATCCAGATTTGGAGTCTCTGACAGCAATCAAT 25927
Qy 1736 CCGCATTTGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAGCGATAAAGATGTA 1795
Db 25928 CAGCTCTCATGACAACTGCTTATGTTTATGATCAAGGTCAGGACATATAGATCTTACGATGCAT 25987
Qy 1796 ACRAACAGGCGGT-----GTGTTTGGATTTGGAGCAGGCGCATGTGAATCCGCAAGG 1849
Db 25988 CAGGAGCAGCTCTTCTCATCGCTTATGATCAAGGTCAGGACATATAGATCTTAAAGAG 26047
Qy 1850 CGATAAACCCGGGATTTGTTTACAACTTCAACAGTGAATGATCAATCACTTACCTTGCA 1909
Db 26048 CTACAGATCTCTGTTTGGTCTACGACATTTGACCTCAAGAGTATTTTGAATTCCTCTGCA 26107
Qy 1910 CTCCTGATTCACAGATCAGATATTTTAGCAATCACTCAATGAACGTCAGCTGCAATG 1969
Db 26108 CTCAGATTTAAGTCCATCAGCTTAAGTATTCACAAACATTCACAAACAGAACCTTGCA 26167
Qy 1970 GAATATTCCGGAACAAACCCGGGTTTGTCTCAATTTACCTCCGATCGATGCGGTATTTCA 2029
Db 26168 AACACACTTCTGCGAAGATCCGGGAACTTGAATCTACCGGCGATATCAGCTTGTCTCC 26227
Qy 2030 AACCTGGCA-----AGACTACGGAGATGATCAAGCGGTGTCACTAACCTTGGGAGTC 2083
Db 26228 CAGAGAACACACATGTTTAAAGCTATGACACTTAGAAGAACAGTACCAATTTGTGCTCTC 26287
Qy 2084 CTAACTCGATATCTCAGTGAATGTCAAGGCTCAGAGGGGATCAAGTATTTGTCATC 2143
Db 26288 ACATTTCAAGTACAAGGTTTCTGCTCGCAATTTCAAGGCGCATCCGTAATCTGCCAGC 26347
Qy 2144 CTAAAGACTTGTGTTCAACACGCTGGATCAGACGCTGAGCTATAGAGT 2192
Db 26348 CCAAAACACTCAACTTCACTTCGGAAGACCAAGAGCTTTCCTACAGGT 26396

RESULT 15

AF360129 2525 bp mRNA linear PLN 18-SEP-2002
LOCUS Arabidopsis thaliana putative subtilisin serine protease
DEFINITION (At3g14240) mRNA, complete cds.
ACCESSION AF360129
VERSION AF360129.1 GI:13430433
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 2525)
Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,

Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Full Length cDNA Clones
Unpublished
2. (bases 1 to 2525)

TITLE
JOURNAL
REFERENCE
AUTHORS

Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Direct Submission
Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

TITLE
JOURNAL

COMMENT
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-length cDNA'). Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X.,
Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D.,
Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Chen, H.,
Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J.,
Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGSC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to Genbank.

FEATURES
source

Location/Qualifiers

1..2525

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="3"

/clone="RAFL09-15-B12 (R09108)"

/note="This clone is in a modified plusuescript vector
(FLC-1) as a BamHI/XhoI insert."

ecotype: Columbia

1..2525

/gene="At3g14240"

1..38

/gene="At3g14240"

39..2366

/gene="At3g14240"

/note="subtilisin proteinase-like"

/codon_start=1

/evidence=experimental

/product="putative subtilisin serine protease"

/protein_id="AAK25839.1"

/db_xref="GI:13430434"

/translation="MARFSYFFLLTLSSSSSSSSNSLTVVVDHAKPSIFPTH
FHWTSLSASITSPFPIIHTVDFHGFSAFLTSQASQLDHPHVISVPEQVRHL
HTTRSPFLGRISDQKGLLESDFGDLVIGDVTGWPEPSFDDRLGVPFIKWK
QCQASQDFPESACNRKLVGRFPGGVVATNGKNETTFRSPRSDGHTHTASIS
AGRVFPASTIGYAGVAAGVAPKRLAAVYCWNSCYDSDILAAPDTAVAGVDVI
SLSGVGVVYPLDALAIGAFADRGIFVYSAGNGPGALFTVNPVPMWTVVAGT
IDRFPANLGNKMGISGVSPGLDPGRMPLVIGSLGGDGISSSLCEGL
DPIVLKGIIVLCRIGNSRATKBEIVRKNGLGMIANGVDFGELVADCHVLPSTV
GASGDEIRRYISSSSKRSKSHPTATVIFKGLRIGPAPVVASFSPGPNPTEPI
LKPDVIAQDLILAAAPDRIGSPGTSNRTBNILSGTSMACPVSGVGLAALLKAAH
PDWSPAAIRSLALITATVDNSGPMWDESGTSSVMDYCSGHVHTKAMDPLVVD
ITSVDYINFLCNSYTRNIVITERQADCDGARRAGHGNLNYSPSVVFPQVGSK
MSTHFRITVNGSDSVYEIKIRPPRTTIVPEKLSFRVQGKLSFVRVKTTEV
KLSFGAUNTVETGHTVMSDGRNVTSPVLTQLQPL"

52

/gene="At3g14240"

/note="compared to genomic sequence resulting in an amino
acid sequence difference"

/replace="t"

2367..2525

/gene="At3g14240"

misc_difference 2508

/gene="At3g14240"

/note="compared to genomic sequence"

/replace="a"

BASE COUNT 585 a 584 c 623 g 733 t

ORIGIN

Query Match

Best Local Similarity 18.2%; Score 424.4; DB 8; Length 2525;

Matches 1075; Conservative 0; Mismatches 916; Indels 51; Gaps 5;

QY 200 CTTCTTCTCGAGCTTCTTACTCTCTATGCTGCTGCGATTGAAGGATTGTGTCAGTTGCA 259

Db 211 CTCTCCCTCCATTTATCCACATACGACACCGTTTTCATGGCTTCTCGGTAGACTCA 270

QY 260 CTGAATCGAAGCCGAGATCTAGATATTCACCTGAAAGTTGTTGCAGTGAGACTGACC 319

Db 271 CTTCAAGACGCTAGTCAACTCTTTAGACCATCTCTCATGTCATCTGTATCCCGGAAC 330

QY 320 ATGTTCTTCAGGTTCAAAACACTTACTCTTACAGTTCTTGGGACTCGACGGTTTGGAA 379

Db 331 AAGTCGCTCACTTGCACACCACTCGTCCCTGAGTTTCTTGGTCTTAGGTCCACCGACA 390

QY 380 ACTCGGCTGATGTTCTAAATCTCGGTTTGGTCAAGGCACCAATATCCGGTGTGATA 439

Db 391 AAGCGGTTCTCTGAAGAGTCTGATTTCCGGTCCGATCTAGTTATCGAGTTATCGATA 450

QY 440 CTGGAGTTTGGCTGAAAGTCTCTAGCTTTGACGATACCGGAATGCTTCGATTCACCGGA 499

Db 451 CTGTGTTTGGCCCGAAGACCTAGCTTTGATGACCGTGGTCTTGGTCTGTTCCTCATTA 510

QY 500 AATGGAAGGATTTGCCAAGAGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGAAGC 559

Db 511 AATGGAAGGCCAATGTATCGCTTCCCAAGATTTTCCGAGTCTGTTGTAACCGTAAAC 570

QY 560 TAATCGGTCTAGATTTCTTCATCAGAGGACACCGTGTGCTTAATTCACAGAGGAATCAC 619

Db 571 TGTGCGAGCTAGATTTCTTCGCGGGGTA-----TGAAGCAACCAACCGGAAAA 621

QY 620 CAAACATGCTCTGTAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCG 679

Db 622 TGAATGAACGACTGAGTTTCCTCTCCGCTGACTCCGATGACATGGGACTCACACAG 681

QY 680 CCTCAACAGTTGGTGGATCTCTGTTTCGATGGCGAATGTTCTTGGCAATGGAGCTGTG 739

Db 682 CTTCAATCTCCGCGCGCGTTACGTTTTTCCGGGCTCAACTCTCGGCTACGCTCAGCGG 741

QY 740 TGGCTCGTGGAGTGGCTCTCGAGCTCACATTCGAGTCTATAAAGTCTGTGTTGTTCAATG 799

Db 742 TCGCTGCTGGGATGGCTCCAAAAGCTAGACTCCGCGCTTACAAAGTCTGTGTTGAACTCCG 801

QY 800 GTTGTGTACAGCTCTGACATTTCTAGAGCTATAGATGTAGCGATTCAAGATAAAGTCGATG 859

Db 802 GTTGTGTACAGCTCCGATATCTTTAGCCGCTTTTCGACACCGCTTTCGCGAGGTGTCGATG 861

QY 860 TTCTTTTCGCTTTCCCTTTGCGGTTTCCCTATTCCCTTTGATGATGACACAAATCGCCATTG 919

Db 862 TCATCTCTCTCCGTTGGAGCGCTGCTGTTCTTATTACCTAGACGCTATCGCTATTG 921

QY 920 GAACATTCGAGCCATGGAAACCGGTATATCTGTAAATCTGTGCACTGGTAAACACGGTTC 979

Db 922 GAGCTTTTGGAGCTATTGACAGAGCATATTCTCTGCTTCCGCGCTTAACCGAGGTTC 981

QY 980 CAAATCGAAGCTCTGTTGCAACACAGCTCTTGGGTCTCAACCATTTGGCGCAGCACCGC 1039

Db 982 CCGGTGCTTTAAACGGTGAACAAATGTTGCTCCGTGGATGACACAGCTCGGAGCTGAACAA 1041

QY 1040 TTGATCGAAGATTTCCCGCTGTGGTCAAGATTAGCCAAACGGAAGAGCTTCTCTATGGAGAT 1099

misc_difference 52

Db 1042 TCGATAGGATTTTCGGCCAAATGTGAACTCGGTAAACGGAGAGATGATTTCTGTGTGTTA 1101
Qy 1100 CATGTATCCGGGAAAGGTATAAA-----GAATCCGGAGAGAGGTGAGGTGA 1150
Db 1102 GTGTATACGGTGGACCGGGTCTGGATCCGGGTGGAATGTACCCGCTTGTATTACGGTGGTA 1161
Qy 1151 TTTACGTCAACAGAGGAGATAAAGGAAGTGAAGTTCCTGTGTTTGAAGAGGTCACTTCCAAGAG 1210
Db 1162 GTTTACTAGCGCGGATGGTTACTCATCGTCTCTGTCTTGAAGGCTCGTTGGATCGGA 1221
Qy 1211 AAGAAATCCGAGGCAAAATGGTGAATTTGATCCGGAGTCAATGGAAGATCGGAAAG 1270
Db 1222 ATTTAGTGAAGGAAGATCGTCTTGTGATAGGAATCAATCTAGAGCAACCAAG 1281
Qy 1271 GAGAGCGGTAAAGAGCTGGAGGATTCGATGATCTTAGCCATACAGAGATCAACC 1330
Db 1282 GTGAGATCGTACGGAATAATGGAGGCTTGGGATGATATAGCAATGGTGTGTTGAG 1341
Qy 1331 AAGAGAAGATCTATTGACGTTTCATCTTACCACTACATGATTTGGTTACACTGAGT 1390
Db 1342 GTGAAGTTTAGTAGCTGATGCGACGTTTACCGGGACATCTGTGTGTTCTGGAG 1401
Qy 1391 CAGTCTTCTGAGGCTTATGTTAATGCCACGGTGA-----CCAA 1432
Db 1402 GAGATGAGATTCGTAGGTATATCTCGAATCATCCAAATCTCGTTCATCGAAACATCCAA 1461
Qy 1433 AGCGCGGATAATTTTGTGTGATCGGTGATTTGGAGGTCAAGAGCACCGAGGTGGCTC 1492
Db 1462 CGGCTACGATGTTTTCAAAAGGTACCGGACTTGGGATCCGACCTGCAACCGGTTGTTGCAT 1521
Qy 1493 AGTTTTCAGTCCGAGGACCGAGTTTACCAATCTTTCGATCTAAACCGGATATCATG 1552
Db 1522 CTTTCTCTGCTGCTGCTTAACTCCAGAGACCGCGGAGATCTTAAACCGGATGTAATCG 1581
Qy 1553 CTCGGGAGTCAATCATTCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCCCTT 1612
Db 1582 CACTGTGTTGAATTTCTAGTCTTGGCTGACCGATTGTCCTCATCTGGTGTACTT 1641
Qy 1613 ATGATTCAAGAGAGTTAACTTCACTGTAATGTAGGAACTTCAATGTCTTGTCCACATG 1672
Db 1642 CTGATAACCGGAGAACTGAGTCAACATTTTATCAGGCACTTCGATGGGTGCGCCGACG 1701
Qy 1673 TTACGGAAATCACTGCTCTTATCCGCTCTGCATACCGGACTGGTCTCCAGCTGCAATCA 1732
Db 1702 TGTCTGTCTAGCTGCTTGTGTTAAGCGGCTCATCCGATGGAGTCCGGCGCGGATAA 1761
Qy 1733 AATCCGATTTGATCAACACAGCGGATTTGTACGATCGTCAAGGAAAGCATAAAGGATG 1792
Db 1762 GATCAGCATTTGATTAACACACGCTTACAGGTTGATTAACAGCGGTGAGCCGATGATGATG 1821
Qy 1793 -----GTACAAACCGCGGTTGTGTTGCGATTGGAGAGGCGATGTGAATCCGAAA 1846
Db 1822 AGTCCACTGGCAATACATCTTCGGTTATGGATTATGGTTCGGGTCTATGTTCCACCAACCA 1881
Qy 1847 AGCGGATAACCGCGGATTTGTTTAAACATTAACAGTGGATTACATAACTTACTCT 1906
Db 1882 AAGCTATGGATCCGGGATTAGTCTATGATATAACATCTTATGATTATATCAACTTCTGT 1941
Qy 1907 GCATCTTGGATTACAGATACATATTTTAGCAATCACTCATAGAACGTTGAGTGA 1966
Db 1942 GTAAATCTTAACACATAGAACCAACATTTGACAAATACCCGTGCGCAAGCGGACTGTG 2001
Qy 1967 ATGGAATATTGCGGAAACCCGGGTTTTAGTCTCAATTAACCGTCTGATAGCCGTGATTT 2026
Db 2002 ACGGTGCAAGACGAGCGGACACCTCGGGAATTTAACTACCCGTCTGTTCTGTCTGTG 2061
Qy 2027 TCAA-----ACGTGGCAAGACTACGGAGATGATCAAGCGGTGTCATTAACGTTG 2077
Db 2062 TTCAGCAGTATGGAGAGAGTAAATGTGACACATTTTATTAGGACTGTGACAAATGTAG 2121
Qy 2078 GGAGTCTTAACATCACTATCTAGTGAATGTCAAGGCTCCAGAGGATCAAGGTTATTG 2137

Db 2122 GCGATTCCGATTTCGGTTTACGAGTTAAGATTAGGCCCTAGAGGACTACGGTGACGG 2181
Qy 2138 TCAATCCTAAAGAGACTTGTGTTCAAACACGCTGGATCAGACGCTAGCTATAGAGTATGGT 2197
Db 2182 TTGAGCCGGAGAGCTATCGTTTCAAGCGGTGGGGCAGAACTGAGTTTGTGTTAGGG 2241
Qy 2198 TT 2199
Db 2242 TT 2243

Search completed: February 2, 2004, 23:28:56
Job time : 5720 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 00:26:16 ; Search time 5004 Seconds
(without alignments)
6335.923 Million cell updates/sec

Title: US-09-806-767-2
Perfect score: 4018
Sequence: 1 MEKPFPLCHIFLLFCSSSS.....NSHLMQVRSPISVTLKTN 775

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_epool/US09806767/runat_02022004_085958_17897/app_query.fasta_1.967
-DB=GenEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09806767.ecgn_1_1_3296 @runat_02022004_085958_17897 -NCPU=6 -ICPU=3
-NO.MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ri:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4018	100.0	78259	8	F20D22 Arabidops
2	2439	60.7	150503	8	AC122149 Cryza sat
3	1742	43.4	22965	6	AX506370 Sequence
4	1742	43.4	22965	8	AY142613 Arabidops
5	1742	43.4	2541	8	AY035090 Arabidops
6	1706.5	42.5	154863	2	OSJN00071 Arabidops
7	1700	42.3	83906	8	AB022220 Arabidops
8	1698	42.3	2525	8	AF360129 Arabidops
9	1697.5	42.2	121160	8	AC005970 Arabidops
10	1684	41.9	2638	8	AY084387 Arabidops
11	1679	41.8	2650	8	LESB42 Lycopersi
12	1679	41.8	5197	8	LES6379 Lycopersi
13	1666.5	41.5	2305	8	AY142612 Arabidops
14	1666.5	41.5	2618	8	AF360285 Arabidops
15	1666.5	41.5	2686	8	AF065639 Arabidops
16	1666.5	41.5	72698	8	AB007645 Arabidops
17	1662.5	41.4	2645	8	AF436834 Arabidops
18	1656	41.2	128017	8	AC084282 Arabidops
19	1653.5	41.2	2274	8	BT001082 Arabidops
20	1653.5	41.2	2520	8	AY091773 Arabidops
21	1642.5	40.9	2549	8	AY080831 Arabidops
22	1642.5	40.9	86212	8	AB010074 Arabidops
23	1640.5	40.8	2423	8	ATARA12 Arabidops
24	1611	40.1	2689	8	LESB41 Lycopersi
25	1611	40.1	5089	8	LES6378 Lycopersi
26	1600	39.8	2390	6	AX654166 Sequence
27	1600	39.8	171099	2	AP003871 Cryza sat
28	1593	39.6	145735	2	AC109601 Cryza sat
29	1589	39.5	2295	6	AX506491 Sequence
30	1589	39.5	2326	8	AY096357 Arabidops
31	1589	39.5	2665	8	AY074375 Arabidops
32	1589	39.5	80346	8	ATM4813 Arabidops
33	1589	39.5	103150	8	ATF1111 Arabidops
34	1589	39.5	195165	8	ATCHRIV82 Arabidops
35	1572.5	39.1	2847	8	LEAJ517 Lycopersi
36	1565	38.9	199414	2	AC124955 Medicago
37	1562	38.9	2465	8	BT002840 Arabidops
38	1554	38.7	113049	8	AC134239 Cryza sat
39	1547.5	38.5	172769	2	AP005570 Cryza sat
40	1542	38.4	2680	8	LEAJ5172 Lycopersi
41	1534.5	38.2	2407	8	LESUBT11 Lycopersi
42	1534.5	38.2	3076	8	LES17276 Lycopersi
43	1521	37.9	2883	8	LES6786 Lycopersi
44	1513.5	37.7	2849	8	LEAJ5173 Lycopersi
45	1509.5	37.6	3173	8	LES17278 Lycopersi

ALIGNMENTS

RESULT 1

```

F20D22
LOCUS       78259 bp      DNA      linear      PLN 20-MAY-1998
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F20D22 sequence, complete
sequence.
ACCESSION   AC002411
VERSION     AC002411.1  GI:2570223
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 78259)
Yu S.K., Ooi, V.S., Osborne, B.I., Schwartz, J.R., Tortumi, M., Kwan, A.,
Yu G., Ooi, V.S., Li, J., Hoang, L., Araujo, R., Au, M.,
Brendel, V., Buehler, E., Conway, A.B., Conway, A.R., Dewar, K.,
Feng, J., Kim, C., Kurtz, D., Li, Y., Palm, C.J., Shinn, P., Sun, H.,
Davis, R.W., Ecker, J.R., Federspiel, N.A. and Theologis, A.
Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence
Unpublished (1998)
2 (bases 1 to 78259)
Theologis, A.
Direct Submission
Submitted (12-AUG-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
3 (bases 1 to 78259)
Theologis, A.
Direct Submission
Submitted (28-OCT-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
4 (bases 1 to 78259)
Theologis, A.
Direct Submission
Submitted (20-MAY-1999) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA
COMMENT     On Oct 28, 1997 this sequence version replaced gi:2323245.
The sequence of BAC F20D22 from Arabidopsis thaliana chromosome 1.
This sequence is of BAC F20D22 from Arabidopsis thaliana chromosome
1. The sequence does not represent the sequence of the entire
insert of this clone. It is shorter by 3823 bp because we sequence
overlapping section only once. In order to facilitate the joining
of overlapping clones in the future for creation of larger contigs,
we provide a small overlap (200 bp) between overlapping submitted
clones. The 5' end of this sequence overlaps by 200 bp the 3' end
of the sequence of the BAC F19P19.
FEATURES             Location/Qualifiers
     1..78259
        /organism="Arabidopsis thaliana"
        /mol_type="genomic DNA"
        /cultivar="Columbia"
        /db_xref="taxon:3702"
        /chromosome="1"
        /clone="F20D22"
        /date="3994..6024"
        /gene="F20D22.1"
        /join(3994..4161,5206..6024)
        /gene="F20D22.1"
        /note="Strong similarity to beta-keto-CoA synthase
        gb|U37088 from Simmondsia chinensis."
        /codon_start=1
        /evidence=not experimental
        /protein_id="AAC16740.1"
        /db_xref="GI:3142289"
        /translation="MNNENHIQSDHMMNTHTVNTKKLPNLLSVRLKYLKGYHLYLSN
        AVVILPVLGLAATSSPSLDTLLVNLHLLKFFHLSLTPAALLILFTLLYFTTRP
        RRIIFLDPRACKPDSLSLITRETTPMDRSQVGFIEDNLAFQOKILERSGLQKTYFP
        EARLRVFPMEARKEATENFVGAIDAVLEKTVGNPKDIGILVNCSLFNPFTSL
        AMTVNKYKGNVLSNLGWMGCSAGLISIDAKQLQVQPNISALVSTENITLWY
        LGNDRMGLNCFIRMGGAVALLSNDRCSKYLQIHTVTRHSGSDNFAFCVYQR
        EDNDNKQGLSVSKNLMAIAGEALKNTITLGLVLPMSQLLFFATILARKVFNVK
        KIKPYIPDFKLAEPFCIHAGGAVLDETEKNLIDLSNEMHPSRMTLAFNGTSSSL
        WYELAYSEAKGRIKRGDRTWQIAFGSGFKCNASVRAALRTIDPSKEKKKTNPWIDEI
        HEFPVPVPTSPVTSSESR"

complement(6494..11181)
/gene="F20D22.2"
complement(join(6494..6724,6808..6909,6996..7346,
7430..7654,7732..7888,8121..8302,8395..8824,8916..9651,
9777..9939,10056..10320,10685..11181))
/gene="F20D22.2"
/note="Contains similarity to adenylate cyclase
gb|AF012921 from Magnaporthe griseae. Est gb|Z24512 comes
from this gene."
/codon_start=1
/evidence=not experimental
/protein_id="AAC16742.1"
/db_xref="GI:3142291"
/translation="MDSKIKKPNALFEDADIDGSESDSTISSVLSLEDSDVVDVSGQ
NLEFLSDNVDSVKGLYFNNVNLIPKISIGLGRRLKLFKFSNEDLDFPPELGNLV
NLEVLQVTKISSPGFGLSDKLGKLELTKVPRKSALTLLSISGLKCLTRLSV
CHFSYILPEVOCGLASLYLDLSPFKIKSLNEIGYLSLFLKVAHNLNLSVFL
ALQNLESDEVENNRNLTTHLPDLNLPRLQILNLRYNKLPISYCIPTIQCNFEGY
EENGVTCSMMVMDVFTPYENNVTTPYHKGSHRNPLMGTG:SSISRCFSARKS
KRWRRQYVQFARQERLNRRKWKGEVPEGLSLKMEVETQKGMKVPONTDRGS
VNSCSDNDKLPFEASVITSEBEESLKADVVSONSQCVETQLTSEDNYESCELT
SPSGDAGKTVDYNSSSERKKPNKSKCSEKYLDPKGSCHKLSTDTNLSKYS
SNFCSSTEDSLPGDFDAGDRPFMTLSKYEVKVLPLDSREVILLDRAKEDVDAITLS
ARALVARLKKLNCLTDPDQVSLDNQVASFALFVSDHFGSDRTAIIERTKAVSG
TNTKPFICTCLTGNQDDLAALNKQVSTTAEDAI:SDVCEKSLRSKXNSIVPLG
KLQFGICRRALMLKCDRMPPVPCELVGVLPDFMHPAMNIVPVKGSVWVMVSLI
ACRPHDIRDTHDOEYFCRYIPLNRNLESIRIKEKLEPGGSVSLSTGKGVANSLI
RCKLGSYAVKMTLEVSGASLDDIRTEYTCLEVRILGALKHDCIVELYGHEISS
KMTISENGEHRVQSILMEHKGSGJAGHIKSEAGKHVPMDLALSIARDISGA
LMELSHKDIHRDIKSENILIDLQNSANGBIKLCDFDRAVPDRHSLHSGCIAHVG
IPPNICVGTPTRMWSEVFRAHMQNFYGLEVDIWSFGCLIFELTLQNPYDLSLQ
IHESLQNGKRPKPKKLETLISETBEESTNKLSEVFDLTSDLTMRFLIDVFHQT
EESPSRLNAGDLHEMLSRKXRE"
complement(12002..16201)
/gene="F20D22.3"
complement(join(12002..12091,12313..12370,12485..12576,
12635..12762,12905..13016,13096..13188,13287..13377,
13468..13586,13763..13876,13958..14086,14252..14426,
14485..14572,14649..14894,15058..15250,15291..15577,
15654..15893,16012..16201))
/gene="F20D22.3"
/note="Contains similarity to gb|Z69902 from C. elegans."
/codon_start=1
/evidence=not experimental
/protein_id="AAC16741.1"
/db_xref="GI:3142290"
/translation="MGVPESTPRKTGGDVSAEYLIATFVGEKSPFLASDFWNKLE
LPLSSRWPSDRVQACELFAQSGNYTRHLAKLIHLISWCLQELLQASDQSSLYKKAV
NATYISSVFLKYLIEGKSDSLQELHLSDESEPPVPHGFVMDODIQNVHMSVLTIG
SNEVRYANADSLYHLFLICFSIRCFILLIILYLAESPXYLHQELNFMMLWTMSQL
LSGSPSGPTDANPFIDAMTQILPHINFRHLWFNDFQEKSLVSLVRRLLNYSRHR
RTPPNAKSYMSDGSQGLIERVGSAAALVLLPLNLYLSNCGSKXNPFLABCSLHVLL
ILINYHKSIMSDSMTDKSDSATSESVKSVHFSSDNTFSKALANADVFEDRSVE
GNAHPAGPHVRIIPASLFTDGLGICILVQILNSMLMLCNNAIWQCRFLADGAVILL
XSLQNSDPEKYLVRTDLDLTPLETLNYSKRTSTSNQIYMLVILLISQDSS
FNSIHKMILLPSVPWYKELHLSHOTSLGSLNVIILITVQHNLSKLRDVLQITCLATL
ANMAHAHLULYASQRLVSLFTYLSKXNKSULDTGDKLQSKINLSGEDVGVSDGL
AASLQFTDFDLRLDILNALITFALPNEPIVYAIMHREQVEFPQKHPFRFHELVN
IYTVCCIGFTHREWSVQKLOQITINNCRSGRWGEMKVSNIYTPSKLYNLPFPMFTQL
HFSVEQESHPEERFIPVWQLAFSRGCGFNPDALNLPVPHVPEXIEIEDERGESEEG
KAKVQELNEQRIVDFP"
17320..19384
/gene="F20D22.4"
join(17320..17477,17799..17913,18007..18057,18148..18321,
18430..18609,18688..18807,18887..19020,19240..19384)
/gene="F20D22.4"
/note="Contains similarity to tetraaricopeptide repeat
protein gb|U46571 from home sapiens. Est gb|Z47802 and
gb|Z48402 come from this gene."
/codon_start=1
/evidence=not experimental
/protein_id="AAC16743.1"
/db_xref="GI:3142292"

```


Qy 401 LeuProArgGluGluLeuLeuArgGlyLysMetValIleCysAspArgGlyValAsnGlyArg 420
Db 63293 CTTTCCAAAGAGAGAAATCCGAGGCAAAATGGTGGATTTGGATCGGGAGCTCAATGGAGA 63352
Qy 421 SerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaSerThr 440
Db 63353 TCGGAGAAAGGAGAGCGGTTAAAGAGCTGGAGGAGTTGCAATGATCTTACCAATACA 63412
Qy 441 GluIleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGly 460
Db 63413 GAGATCAACCAAGAGAAGATTCTATTGACGTTCTCTTACCAGCTACATTGATTGT 63472
Qy 461 TyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArg 480
Db 63473 TACACTGAGTCAGTCTCTGAGGCTTATGTTAATGCCACGGTGAACCAAGGCGCG 63532
Qy 481 IleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSer 500
Db 63533 ATAATTTTGGTGATCGGTCGATGGAGGTCACGAGCACCGAGGTCGCTCAGTTTCA 63592
Qy 501 AlaArgGlyProSerIleAlaAsnProSerIleLeuLysProAspMetIleAlaProGly 520
Db 63593 GCTCGAGGACGAGTTAGCAATCTCTGATCTAAACCCGATATGATTGCTCCGGA 63652
Qy 521 ValAsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSer 540
Db 63653 GTCAATATCATTTGGGCTTGGCTTCCAAATCTAGGACCAACCGGACTTCTTATGATCA 63712
Qy 541 ArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGly 560
Db 63713 AGAAGAGTTAACTTCACTGTAATGTGAGGAACTCAATGTCTGTCCACATGTAGCGGA 63772
Qy 561 IleThrAlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaIleLysSerAla 580
Db 63773 ATCACTGCTTTATCCGCTGTCATACCGAACTGTCTCCAGTGCATCAATCAATCCGCA 63832
Qy 581 LeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaTyrLysAspGlyLys 600
Db 63833 TTGATGCAACACGCGAATTTGTACGATCGTCAAGGAAAGCGATAAAGATGTTAAACAA 63892
Qy 601 ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 63893 CCAGCCGGTGTGTTGGATTTGGAGCAGCGCATGTAATCCGAAAGCGGTAATACCCG 63952
Qy 621 GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrIleuCysThrLeuGlyPhe 640
Db 63953 GGAATGGTTTACAACTTCAACCGAGTGGATTACATAACTTACCTCTGCACCTCTGGATTC 64012
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 64013 ACAAGATCAGATATTTAGCAATCACTCATTAAGACGTCAGCTGCAATGGAATATGGCG 64072
Qy 661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 64073 AAAAACCCGGTTTGTAGTCTCAATTCACCTCGATAGCCGTGATTTTCAAGCTGGCAG 64132
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSer 700
Db 64133 ACTACCGAGATGATCAACAGCGTGTCACTTAACGTTGGAGTCTTAACTCGATATACTCA 64192
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db 64193 GTGAATGTCAGGCTCCAGAGGGGATCAAGTTATTTGTCAATCCTTAAGAGACTTGTGTTTC 64252
Qy 721 LysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuLysLysLysAsnArg 740
Db 64253 AAACACCTGGATCAGCGCTAGCTATAGATGATGTTGTTATTAAGAGAAACACAGA 64312
Qy 741 GlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeu 760
Db 64313 GGAGGGAGGTGGCTAGCTTTTGCAAGGCGACTTGACTTGGTCACTCTCATATCTG 64372
Qy 761 MetGlnArgValArgSerProIleSerValThrLeuLysThrAsn 775

Db 64373 ATGCAGCGAGTTAGAAGTCCAATCTCTGTAACTTGAAGACTAAC 64417
RESULT 2
AC122149
LOCUS
DEFINITION
OSUNBa0083D24, complete sequence.
ACCESSION
AC122149.2 GI:29611669
KEYWORDS
HTG.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 150503)
Wing, R.A., Yu, Y., Soderlund, C., Kim, H.-R., Rambo, T., Currie, J.,
Collura, K., McCombie, D.W.R., de la Bastide, M., Spiegel, L.,
Preston, R., Kirchoff, K., Kuit, K., Nascimento, L., Zutavern, T.,
Balijs, V., Bell, M., Baker, J., Santos, L., Miller, B.,
Katzenberger, F., Muller, S., King, L., Yang, C., O'Shaughnessy, A.,
Palmer, L. and Dedhia, N.
Rice Genomic Sequence
Unpublished
2 (bases 1 to 150503)
Wing, R.A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T.,
Saski, C., Henry, D., Gates, R., Simmons, J., McCombie, D.W.R., de la
Bastide, M., Spiegel, L., Preston, R., Kirchoff, K., Kuit, K.,
Nascimento, L., Zutavern, T., Balijs, V., Bell, M., Baker, J.,
Santos, L., Miller, B., Katzenberger, F., Muller, S., King, L., Yang, C.,
O'Shaughnessy, A., Palmer, L. and Dedhia, N.
Direct Submission
Submitted (23-MAY-2002) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
3 (bases 1 to 150503)
Wing, R.A., Yu, Y., Soderlund, C., Kim, H.-R., Rambo, T., Currie, J.,
Collura, K., McCombie, D.W.R., de la Bastide, M., Spiegel, L.,
Preston, R., Kirchoff, K., Kuit, K., Nascimento, L., Zutavern, T.,
Balijs, V., Bell, M., Baker, J., Santos, L., Miller, B.,
Katzenberger, F., Muller, S., King, L., Yang, C., O'Shaughnessy, A.,
Palmer, L. and Dedhia, N.
Direct Submission
Submitted (08-APR-2003) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
On Apr 8, 2003 this sequence version replaced gi:21104882.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by more than
one plasmid subclone; and the assembly was confirmed by
restriction digest. There is only PCR coverage at the following
locations: 23285-24006, 119474-119561 and 121193-121700. At bases
128212-128214 there is only single stranded transposon coverage and
at 121172-122424, 122594-122826, and 123373-123660 there is double
stranded transposon coverage. There is dinucleotide (TA) repeat at
52845-52892 of 15-24 pairs. Syntacta reads are the only coverage
at 57061-57199. There are Bacterial transposons at the following
locations: 628-3459, 26704-28941, 130907-134220, 23095-24280,
118487-121172, 120153-124368, 123584-125272, 128135-130129,
127157-130772, and 130790-133540. The assembly overlaps from base
144347-150503 with OJ1172F09 (accession # AC119796). The overlap is
from 1-6156 bases on OJ1172F09. The nucleotide sequence of this BAC
clone was generated by combining Syngenta, Cold Spring Harbor
Laboratory, and Arizona Genomics Institute sequencing data.
FEATURES
Location/Qualifiers
1. 150503
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39947"
/chromosome="3"
/clone="OSUNBa0083D24"

QY	291	SerLeuGlyGlyPheProLeuTyAspAspThrIleAlaIleGlyThrPheArg	310
QY	291	SerLeuGlyGlyPheProLeuTyAspAspThrIleAlaIleGlyThrPheArg	310
DB	87236	TCCCTCGCGGCTTCCCATCCGCTCTTCGAGGACAGCATGCCATCGGACGCTTCGCT	87295
QY	311	AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSer	330
DB	87296	GCCACGACGATGCGCTTCCGTCGTGTCGCGCGCGGAAACAACGGGCGATCGCGGAC	87355
QY	331	SerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArg	350
DB	87356	TCCGTCGCCAACGAGCGGCATCGGTGATCACGTCGCGCGCGGACGCTGACCGGGA	87415
QY	351	PheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyTrpGlyLysLeuTyTrp	370
DB	87416	TTCCCGGCTTATGTCGCGTCGGAACGCGCGATCTCTGACGGGAGTTCATGTTCC	87475
QY	371	GlyLys---GlyIleLysAsnAlaGlyArgGluValGluValIleTyValThrGly	389
DB	87476	GGGAAGTCATTTGAAATGGCGGGAAGGAGCTCGAGCTGGTGTACGCCCGCAGTGG	87535
QY	390	AspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLys	409
DB	87536	ACCGGGAAGAGATGATCTGCATCAAGGGGCGCTGTCTGCCGACCTGCGCGGAAG	87595
QY	410	MetValleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGlu	429
DB	87596	ATGTCGTGTGTGACCGTGCGCATACCGGACGGGTGACAAAGGCGAGGAGTAAACA	87655
QY	430	AlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIle	449
DB	87656	GCAGCGCGCGCCCATGATCTCGGAATCCGAGATAAACAGAGAGAGATCTCGTC	87715
QY	450	AspValHisLeuLeuProAlaThrLeuIleGlyTyThrGluSerValLeuLeuLysAla	469
DB	87716	GATGTCCATGTCGTCGCTGACGCTCATCGCTACGAGGAGCGCTCGAGTCAAGAAC	87775
QY	470	TyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIleGly	489
DB	87776	TACGTACGTCACGCGCGCGGTGGGAGATAGTGTTCGGGGGACGCGGATGG	87835
QY	490	ArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnPro	509
DB	87836	CGGCGCGCGCGCGGTGGGCTGTCTCCGCGCGCGGCGGAGCTGACGAACCG	87895
QY	510	SerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGln	529
DB	87896	TCCGTGCTCAAGCCGACGTGTGCGCGCGCGCGGTGAAACATCATCGCGCGTCC	87955
QY	530	AsnLeuGlyProThrGlyLeuProTyTrpSerArgValAsnPheThrValMetSer	549
DB	87956	AACCTGGACCGTCGGGGTGGAGGGCGAGCGCGCGCGCTCGAGTTCACGCTCTCG	88015
QY	550	GlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTrp	569
DB	88016	GGGACATCAATGCGTCCGCGACGTCGAGCGCATCGCGCGCTCATCAGTTCGGCGAC	88075
QY	570	ProAsnTrpSerProAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyTrp	589
DB	88076	CCGTCGTGGAGCGCGCATGTCGCGTCCGCGATCATGACGCGCGCATGTCGCGAC	88135
QY	590	ArgGlnGlyLysAlaIleLysAspGlyAsn---LysProAlaGlyValPheAlaIleGly	608
DB	88136	CGCGAGGCGAAGCGCATATGACGCGGAACGCGCGCAAGCGCGCATGCGCGTGGC	88195
QY	609	AlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyTrpAsnIlePro	628
DB	88196	GCGCGGACGTAACCGCGGAGCGCGTTCGACCGCGCGCTTCGATCGATCGACCGC	88255
QY	629	ValAspTyIleThrTyTrpLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIle	648
DB	88256	GCGCATGCTCACCCACCTCTGCAACCTCGGTTACACCCATCGAGATCTTCAAGATC	88315
QY	649	ThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsn	668

Db 1177 AAGATTGTTGTTGATAGAGGTGTTAATGCTAGAGTTGAGAAGAGAGCTGTGTTAGA 1236

Qy 429 GluAlaGlyGlyValAlaMeIleLeuAlaAsnThrGluIleAsnGlnGluAspSer 448

Db 1237 GATGCTGTTGTTAGGATGATAATGCGAATAGTCTGCGAGTGGAGAGAGCTGTGTT 1296

Qy 449 IleAspValHisLeuLeuProAlaThrLeuIleGlyThrGluSerValLeuLeuLys 468

Db 1297 GCGGATAGTCAATTCCTCCCGCATCGCTAGGAGAGAGAGCTGTGTTACTTACCTAGG 1356

Qy 469 AlaTyrValAsnAlaThrValIleProLysAlaArgIlePheGlyGlyThrValIle 488

Db 1357 GAGTAGTGTAGTCAATTCACCAACCGCTCTCTGTTTAAAGAGAACGGTCTT 1416

Qy 489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508

Db 1417 GACGTTAAGCCGCTCCTGCTGTTGCTGTTTACGTCGAGAGTCTAATACCTGTTACT 1476

Qy 509 ProSerIleLeuLysProAspMetIleAlaProGlyValAlaSerIleIleAlaAlaTyrPro 528

Db 1477 CCGTGAATCTTGAAGCCTGATGTTATGCTCCTCGAGTTAATATTTTGGCTGTGTTGCT 1536

Qy 529 GlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMet 548

Db 1537 GACGCTATTGTCCTTACTGCTCTTCAACAGGACTCTAGGAGAGTCTCAGTTCAACATCATG 1596

Qy 549 SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAla 568

Db 1597 TCAGGTAGCTCAATGTCATGCCACACATCAGTGTGTTAGCGGCTCTTTTGAAGCAGCT 1656

Qy 569 TyrProAsnTyrSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyr 588

Db 1657 CACCTGTAGTGGAGTCCGAGTGTATCAATCAGCTCTCATGACTCAGCTTACGTTCTT 1716

Qy 589 AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606

Db 1717 GACACACCAACGCTCTCTCCATGATGCTCAGACACACACGCTTATCAACCATATGCT 1776

Qy 607 IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIle 626

Db 1777 CACGCTCGGCGCATGATGATCCCAAGAGCTCTCCACCGAGTCTTGCTAGCATC 1836

Qy 627 GlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeu 646

Db 1837 TCACCGAGGAATACATCAGGTTTGTGCTCTCTAGACTACACAGTCGATCAGTGT 1896

Qy 647 AlaIleThrHisLys-----AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe 665

Db 1897 GCGATTGTGAAGCAGACCTAGCGTTAACTGCTCGAAGAAGTTC---TCAGATCCTGCT 1950

Qy 666 SerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIle 685

Db 1951 CAGCTCAACTACCAAGTTCTCGGTTTGTGTT---GGGGGTAAGAGAGTGTGCGGTAC 2007

Qy 686 ThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAla 705

Db 2008 ACTCGGAGAGTACAATGTTGTCGACAGCTCGGTTTACAAGTGACGGTTAATGGA 2067

Qy 706 ProGluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGln 725

Db 2068 GCTCTAGTGTGCGAATCTCTGTTAAACCATCGAACTTTCGTTTAAAGCGTGGGAG 2127

Qy 726 ThrLeuSerTyrArgValThrPheValLeuLysLysLysAsnArgGlyGlyLysValAla 745

Db 2128 AAGAGAGAGGTACACAGTCACGTTTGTGTAGCAAGAAGAGAGTGTAGTATGACGAACAAGCT 2187

Qy 746 SerPheAlaGlnGlyGlnLeuThrTyrValAsnSerHisAsnLeuMetGlnArgValArg 765

Db 2188 GAGTTT-----GGTTCCATCCTTGGAGCAATCCGAGCAC-----GAAGTGAGA 2232

Qy 766 SerProIleSer 769

Db 2233 AGTCCCGTTGCA 2244

RESULT 4

AY142613

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .2296

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="2"

/clone="U09230"

/note="This clone is in pJUN1 51."

ecotype: Columbia

1. .2296

/gene="At2G05920"

1. .2265

/gene="At2G05920"

/codon_start=1

/evidence=experimental

/product="putative subtilisin serine protease"

/protein_id="AA113182.1"

/db_xref="GI:23296838"

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

db

AY142613 2296 bp mRNA linear PLN 23-SEP-2002
Arabidopsis thaliana putative subtilisin serine protease
(At2G05920) mRNA, complete cds.

AY142613.1 GI:23296837

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 2296)

Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W.,
Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C.,
Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R.,
Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J.,
Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
Davis,R.W., Ecker,J.R. and Theologis,A.

Arabidopsis Open Reading Frame (ORF) Clones

Unpublished

2 (bases 1 to 2296)

Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W.,
Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K.,
Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R.,
Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J.,
Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
Davis,R.W., Ecker,J.R. and Theologis,A.

Direct Submission

Submitted (16-AUG-2002) Plant Gene Expression Center, 800 Buchanan

Street, Albany, CA 94710, USA

The RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and
sequenced the pJUN1 (ORF) clones using the RAFL cDNAs: Yamada,K.,
Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W., Lee,J.M.,
Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J.,
Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T.,
Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP
/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.

Location/Qualifiers

1. .2296

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="2"

/clone="U09230"

/note="This clone is in pJUN1 51."

ecotype: Columbia

1. .2296

/gene="At2G05920"

1. .2265

/gene="At2G05920"

/codon_start=1

/evidence=experimental

/product="putative subtilisin serine protease"

/protein_id="AA113182.1"

/db_xref="GI:23296838"

gene

CDS

/translation="WASSSSSSSSIIITITFLLELLHTTAKKTYIIIRVHSDKPSFL
 THDWYSQLNSSESLIYITITSPHGSAYLSDSTEADSLSSNSLIDIFEDPLYLTH
 TRTPFELGLNSEFGHDLSSNGVIIGVLDGTVMPESSRSDFTDMPPIPSKWKGE
 ESGSDPFLKCKKLIIGARFSGKQWASGGFSSKRESVSRDVGHGHTSTTAAG
 SAVNAGFLVAACTARMAATRAVATYKVCWSTGCFGSDILAAMDRAILLDGVLSL
 SLGGSPAYEDTITAGAFAMBERGVFVSCAGNSGPTASVANVAWPMWTVGAGTLD
 RDPFAPNLNGKRLTGVLSYSGVGMGKPELELYNKNSSSNLCLPSLSSSIYRG
 KIVCDGVNARVKGAVRDAGGLGMIMANTASGEELVADSHLLPALAVGKKTGDL
 LREYVSDSKPTALLVFKGTVDLYKPSVVAFAFSSRGPNTVPEILKPDVPGVNI
 AGWSDAIGPTGLDKDSRTOFNIMSGTSMSCPHISGLAGLLKAHPWSPSAIKSALM
 TTVAVLNTWAPLHDAANDSLNPNYAHSGHVDPOKALSGLVVDIISTREYIRFLCSL
 DYTVDHIVAIKRVNCSKSPDGLNYPESVLFGKRVVYTRVETVNVGAASV
 KYVYNGAPSGVGSVKPSKLSFKSVGKKRYITVTFVSKGVSNKNAEFGSITWSNPQ
 HEVRSVPFAFNSNRF"
 2266..2296
 /gene="At2g05920"

3' UTR
 BASE COUNT 532 a 515 c 546 g 703 t
 ORIGIN

Alignment Scores:

Pred. No.: 3,54e-104 Length: 2296
 Score: 1742.00 Matches: 370
 Percent Similarity: 63.48% Conservative: 115
 Best Local Similarity: 48.43% Mismatches: 241
 Query Match: 43.35% Indels: 38
 DB: 8 Gaps: 15

US-09-806-767-2 (1-775) x AV142613 (1-2296)

QY 11 IlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLysGlnThrThrIleVal 30
 DB 52 CTCTCTCTCTCTCCACACACACACACGCGC-----AAAAAACCTACATCATC 96
 QY 31 GlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeu 50
 DB 97 CGGTCGAAT---CACTCCGATAAACCGGAATCATCTCTCACTCACACGATTGGTACACA 153
 QY 51 SerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSerArg 70
 DB 154 TCTCAACTCAATTCA-----GNAATCATCT 177
 QY 71 LeuLeuThrSerThrGlySerAlaIleGluGluPheAlaAlaGlnLeuThrGluSerGlu 90
 DB 178 CTCTCTACACTTACACACACTCTCTCCATGGGCTTCTCCGCTTACCTCGACTCCACCGAA 237
 QY 91 AlaGlu---IleLeuArgTyrSerProGluValValAlaValArgProAspHisValLeu 109
 DB 238 CGCGATTCTCTCTCTCTCTCAACTCAATCTCGATATCTTCCGAAGATCTCTCTAC 297
 QY 110 GlnValGlnThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 128
 DB 298 ACATCTCACACTACGCGTACTCTCGATTCTCGGTCTCAATCCGAATCGGTGTTCAAC 357
 QY 129 GlyValTrpSerLysSerArgPheGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 148
 DB 358 GATCTCGGTCTCTCTCT-----AACCGGTTATCATCGAGTTTATGATCTGCG 408
 QY 149 ValTrpProGluSerProSerPheAspAspThrGlyWecProSerIleProArgLysTrp 168
 DB 409 GTTGTGGCTGAATCTAGAGCTTCGATGATACATGATATGCTGAGATTCTCTTCAATG 468
 QY 169 LysGlyIleCysGlnGlnGluGluSerPheSerSerSerSerSerSerSerSerSerSerSer 188
 DB 469 AAGAGAGATGTAATCTGTTCCGATTTCGATTCGAATGTTGTAAACGAAGCTTATC 528
 QY 189 GlyAlaArgPhePheIleArgGlyHisArgValAlaAlaAsnSerProGluGluSerProAsn 208
 DB 529 GGAGCTAGAAGCTTCTCCAAAGGATTTCAATAGCTTCTCGGTGGTGTTCGAGTAAG 588
 QY 209 MetProArgGluThrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer 228
 DB 589 -----CGTGAATCTGTTTCTCTCGTGTGTTGTCGAGCATGGACATGGACATCACTTCAACT 642

QY 229 ThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAla 248
 DB 643 ACCGCCGGGGATCCCGCTTAGAACCGCTTCTCGGTACCGCCCGCGTACGCGC 702
 QY 249 ArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys 268
 DB 703 AGAGTATGGCCACTCGTGTCTGCTTACTTATAAAGTTTGTGGAGTACTGTTGT 762
 QY 269 TyrSerSerAspIleLeuAlaAlaAspValAlaIleGlnAspLysValAspValLeu 288
 DB 763 TTTGGATCTGATATACTAGCTGCTATGATCGAGCTATACTTGTGTTGTTGATGCTGT 822
 QY 289 SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThr 308
 DB 823 TCGTTATCTCTTGTGTGTGTCTCTCGTATATCGGATACGATTCGATTCGATTCGAGCG 882
 QY 309 PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle 328
 DB 883 TTTTCGGCTATGAGAGAGAGGTGTTTGTCTGCTGTGTAATAGTGGTCCCTACT 942
 QY 329 GluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyValAlaGlyThrLeuAsp 348
 DB 943 AGAGCTTCTGTGTCCAAATGTTGCTCTCTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1002
 QY 349 ArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeu 368
 DB 1003 AGAGATTTTCGGCTTTTCGGAATCTCGGTAAACGGGAAACGACTTACCGGTGTTTCGCTG 1062
 QY 369 TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGly 388
 DB 1063 TATACGGCTGTAGGAATG-----GGGACGAAAGCGCTTGAATTTGTTTATAATAAGGG 1116
 QY 389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly 408
 DB 1117 AATAGTAGTTCGAGTAATCTTTGTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1176
 QY 409 LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys 428
 DB 1177 AAGATTGTTGTTGATAGAGGTGTTAATGCTAGAGTTGAGAAAGAGGCTGCTGTAGAGA 1236
 QY 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSer 448
 DB 1237 GATGCTGTGTTTAGGGATGATAATGCGAATACTGCTGCGAGTGGAGAGGAGCTTGTG 1296
 QY 449 IleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLys 468
 DB 1297 CGGATAGTACTATTGCTTCCCGCATCGCTGAGGAGAGAGCTGCTGATTTACTTAGG 1356
 QY 469 AlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIle 488
 DB 1357 GAGTATGTTAAGTCAGATTCTAAACCAACCGCTCTTCTGTTTAAAGGAACGGTCTTT 1416
 QY 489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508
 DB 1417 GACGTTAAGCGCTCTCTGTTGTTGCTGCTTTAGCTCGAGAGCTCTAATATCTGTTACT 1476
 QY 509 ProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaTrpPro 528
 DB 1477 CCTGAAATCTTGAAGCCTGATGTTATTCGCTCGAGCTTAATATTTTGGCTGTTGCTCT 1536
 QY 529 GlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsnPheThrValMet 548
 DB 1537 GACCTATTGTTCTCTGTTGTTGACAGACTCTTAGGAGGACTCAGTTCAACATCATG 1596
 QY 549 SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAla 568
 DB 1597 TCAGTACGTCATGTCATGCCACACATCAGTGGTTTACGGGTCTTTTGAAGCAGCT 1656
 QY 569 TyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyr 588
 DB 1657 CACCTGAGTGGAGTCCGAGTGTCTATCAATCAGCTCTCATGACTACAGCTTACGTTCTTT 1716
 QY 589 AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606


```

Db 1717 GACACACCAACGCTCTCTCCATGATGCTGCGACACAGCAGCTATCTACCACTATGCT 1776
Qy 607 IledyAlaGlyHisValAsnProGlnLeuAlaLeAsnProGlyLeuValTyAsnIle 626
Db 1777 CACGGCTCGGCCATGATGATCCCAAAAGGCTCTCTACCAAGGCTCTGCTACGACATC 1836
Qy 627 GlnProValAspTyIleThrTyLeuCyethrLeuGlyPheThrArgSerLeuLeu 646
Db 1837 TCACACGAGGATACATCAGGCTTTTGTGCTCTCTAGACTACACAGCTCGCATCTGTT 1896
Qy 647 AlaIleThrHisIlys---AsnValSerCysAsnGlyIleLeuArgIlyAsnProGlyPhe 665
Db 1897 GCGATTCTGAAGCGACCTAGCGTTAACTGCTCGAAGAAGTTC---TCAGATCTCTGCT 1950
Qy 666 SerLeuAsnTyProSerIleAlaValIlePheIysArgGlyIysThrGluMetIle 685
Db 1951 CAGCTCACTACCAAGTTCTCGGTTTCTTT---GGGGTAAAGAGTTGTCGGGTAC 2007
Qy 686 ThrArgValThrAsnValGlySerProAsnSerIleTySerValAsnValIysAla 705
Db 2008 ACTCGGAAGTAAACAATGTTGTGTCGACGAAGCTCGGTTTACAAAGTGACGGTTAATGGA 2067
Qy 706 ProGluGlyIleIysValIleValAsnProIysArgIleValPheIysHisValAspGln 725
Db 2068 GCTCTAGTGTGGAATCTCTGTAAACCATCGAAACTTCGTTTAAAGCGTGGAGAG 2127
Qy 726 ThrLeuSerTyArgValTrpPheValLeuIysIysAsnArgGlyGlyIysValAla 745
Db 2128 AAGNAGAGGTACACAGTCAGCTTTGTAGCAAGAAAGGAGTGTAGTACGACAAAGGCT 2187
Qy 746 SerPheAlaGlnGlyLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GGTTCATCACTTGGAGCAATCCGACGAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTGTCA 2244

RESULT 5
AY035090
LOCUS AY035090 2541 bp mRNA linear PLN 18-SEP-2002
DEFINITION Arabidopsis thaliana putative subtilisin serine protease
            (At2g05920) mRNA, complete cds.
ACCESSION AY035090
VERSION AY035090.1 GI:14334833
KEYWORDS FULL CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 2541)
AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
Davis,R.W., Ecker,J.R. and Theologis,A.
TITLE Arabidopsis Full Length CDNA Clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2541)
AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G.,
Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and
Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2001) Plant Gene Expression Center, 800 Buchanan

```

Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L.,
Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Koesema,E.,
Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J.,
Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGECC) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.

Location/Qualifiers
source
1. .2541
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAFL09-15-A05 (R09230)"
/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert."
ecotype: Columbia
gene
1. .2541
/gene="At2g05920"
1. .22
/gene="At2g05920"
misc_difference
2
/gene="At2g05920"
/note="compared to genomic sequence"
/replace="C"
23. .2287
/gene="At2g05920"
/codon_start=1
/evidence=experimental
/product="putative subtilisin serine protease"
/protein_id="AAK59595.1"
/db_xref="GI:14334834"
/translation="MASSSSSSSITITFTFLHLLHTAKTYIIRVNSDKPESPL
THDWTYQLNSSESLYTYTTFHGFSAVLDSTEADSLSSNSILDIPEDPLYTLH
TRTPPEFLGNSPFGVHDLGSSNGVIGVLDGVMPESRFDPTDMPETPSKWKGC
ESGSDPDKLCKNKLIGARSPGFQMGAGGFGSSKREVSFPRVDGHGTHTTTAAG
SAVNASPLGYAAGTARGMATRVATYKVCWSTGCGSDILAAMDRAILDGVVLSL
SLGGSAPYTRDTIAGFAMERGVFVSCAGNSGPTRASVANVAPWVMTVAGIJD
RDPPAFANLNGKRLTGVSILYSGVGMGTKELELVYKNGSSNSLCLPGSLDSSIVRG
KIVVCDGVNARVEKGVADAGGLGMIMANTAAAGELVADSHLLIPAIAGKKTGDL
LREVKSDKPTALLVFKGLDVPSPVAAFPSSGPNVTTPILKPDVPGVNTL
AGHSDAIGTGLDKDQSPRTQFNMSGTSNCPHISGLAGLLKAHPEWSPALKSALM
TTAYLVNTNAPLHADNLSNPYAGSHGVDPQALSPGLYDLSIEYIPLCSL
DYTVDHVAVLKPSVNCSSKSKFSGVGLQNTYPSFVLFQGVKRVYTRVNTVGAASSV
YKVTNAGASGVGISVSKPSKLSFKSVGKEKTYTTFVSKGVSMTNKAEGFSTWSNPQ
HEVKSVPVAFSWNR"

3'UTR
2288..2541
/gene="At2g05920"
misc_difference
2524..2525
/gene="At2g05920"
/note="not present in genomic sequence"
BASE COUNT 617 a 549 c 600 g
ORIGIN

Alignment Scores:
Pred. No.: 4.07e-104 Length: 2541
Score: 1742.00 Matches: 370
Percent Similarity: 63.48% Conservative: 115
Best Local Similarity: 48.43% Mismatches: 241

Qy 266 AsnGlyCysTyrSerAspIleLeuAlaAlaAspValAlaIleGlnAspLysVal 285
 Db 65839 GGAGGGTGTCTAGCTCCGACATCTGGCGCGCTGATCGCGCGTGTGGACGGCTC 65780
 Qy 286 AspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAla 305
 Db 65779 GAGTGTCTCTCCATCTCTCCCTGGGGTGGCGCTCCCGTACTACCTCGACAGCTGTCC 65720
 Qy 306 IleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsn 325
 Db 65719 ATAGCTGTCTGGGGCATCGAGATGGGCTGTCTGCTGCTCGCGCGGCAACGCC 65660
 Qy 326 GlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGly 345
 Db 65659 GGGCTACCCGATAGCCCTCACCACTGTGCGCGTGGATACCAACGAGTGGGCGGAGC 65600
 Qy 346 ThrLeuAspArgPheProAlaValValArgLeuAlaAsnGlyLysLeuTyrGly 365
 Db 65599 ACCATGACCGTGTCTCCGCGGACGCTCGCAACGCGCGCAACATCATCCGCG 65540
 Qy 366 GluSerLeuTyrProGlyLysGlyIleLysAsnAlaGly-----ArgGluValGluVal 383
 Db 65539 GTCTCGCTTAC-----AAGCGCTCGGAATCTGTGCGCGGACGAGCATCCGGTG 65486
 Qy 384 IleTyrValThrGlyGlyAspLysGlySerGlu-----PheCysLeuArgGlySer 400
 Db 65485 GTCTACTTGGCGCAATTCAGCATCGCGACCCCGCTCTCTGTGCTCGCGGGAGCG 65426
 Qy 401 LeuProArgGluLurIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArg 420
 Db 65425 CTGACGCGCACGACGCTCCGCGAAGATAGTGTGACCGCGGCATCATGCTCCG 65366
 Qy 421 SerGluLysGlyGluAlaValLysGlyAlaGlyValAlaMetIleLeuAlaAsnThr 440
 Db 65365 GTGCAAGAGGTGAGTGTTCAGAGGAGCGCGCATCGCATATCTTCCCAACACC 65306
 Qy 441 GluIleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGly 460
 Db 65305 CGCGCAACCGCGAGGAGTGTGCGCGACGACCTGTGCGCGCGCTGCGAGTCCGC 65246
 Qy 461 TyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArg 480
 Db 65245 GAGCGGAAGGATTCGCGCAAGAGCTACAGCAAGAGCGCCCAAAACCGCGGAGC 65186
 Qy 481 IleIlePheGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSer 500
 Db 65185 CTCAGCTTCGCGGAACCAATTCGCGCATCCGCGCTGCGCGGTGTGCTCGCGGTCTCG 65126
 Qy 501 AlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGly 520
 Db 65125 TCCGCGGACCAACATCTGTGATCTGTGAGATCTCAAGCGGACGTGTCGCGCGCGC 65066
 Qy 521 ValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSer 540
 Db 65065 GTGACATCTTGGCGGATGAGCGCGGACCGCGCGCTGCGAGTGTTCAGCGACAC 65006
 Qy 541 ArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGly 560
 Db 65005 CGTCCGCTGGGTTCACATCTGTGCGGACGTCCATGTCGCGCGCGACGTGCGCGC 64946
 Qy 561 IleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAla 580
 Db 64945 GTGCGCGCTGTATCAAGCGGAGCCACCGGATGAGGCGCCACACAGATCAAGTCCGG 64886
 Qy 581 LeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGly----- 598
 Db 64885 CTGATGACCGCGGTACTGTCAGCAACACAGTACGCGCGCATGAGAGCGCGGAC 64826
 Qy 599 AsnLysProAlaGlyValPheAlaIleGlyValAlaGlyHisValAsnProGlnLysAlaIle 618
 Db 64825 GGCAAGCGTTCGAGCGGCTTCGAGCAGCGAGCGGCGCACATACACCGGTGCGAGTCTC 64766
 Qy 619 AsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeu 638

Db 64765 ACCCTGCGCTGTCTAGCATTTGGCCAGCCGACTACCTGGAGTTCTCTCTGCGACGAG 64706
 Qy 639 GlyPheThrArgSerAspIleLeuAlaIleThr-----HisLysAsnValSerCysAsnGly 657
 Db 64705 CACATGACCGCGATGCGAGCTCAGGACCTTCACCAAGAACTCCACATGATGAGCGAC 64646
 Qy 658 IleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLys 677
 Db 64645 ACCTTC-----AGTTCGCGCGGACCTCACTATCTCGGCCATCTCAGTCTCTCTGCC 64592
 Qy 678 ArgGlyLysThrThrGluMetIleThrArgArg-----ValThrAsnValGlySerProAsn 696
 Db 64591 CACCAAGCATCAAGGCGCTGACGTCGCGGCTGTCACGACAGCTGTGCTCTCTCTCTCT 64532
 Qy 697 SerIleTyrSerValAsnValLysAlaProGluGlyLysValIleValAsnProLys 716
 Db 64531 TCGACGATACCATGTCACAGGTCACCAAGTCAAGGCGCGCGATGTCATCGTCAACCGAAC 64472
 Qy 717 ArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuLys 736
 Db 64471 ACCTGTCATCTGTAAGCAACACCAAGCTTAGCTTAGGTTACAGTACAGTACACCAAG 64412
 Qy 737 LysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsn 756
 Db 64411 GCCGCCAG-----AAGGACCGGAGTTC-----GGAGCCTGTCTCTGAGCGAC 64367
 Qy 757 SerHisAsnLeuMetGlnArgValArgSerProIleSerValThr 771
 Db 64366 GGCCTCCACAT-----GTCAGAGCGCTGTCTGCTCTCTCTCTCTCTCTCTCT 64331
 RESULT 7
 AB022220/c AB022220 83906 bp DNA linear PLN 27-DEC-2000
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MLN21.
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MLN21.
 ACCESSION AB022220 BA000014
 VERSION AB022220.1 GI:4159709
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi-
 1 (sites)
 Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
 Structural analysis of Arabidopsis thaliana chromosome 3. I.
 Sequence features of the regions of 4,504,864 bp covered by sixty
 pl and TAC clones
 DNA Res. 7 (2), 131-135 (2000)
 JOURNAL 20277480
 MEDLINE 10819329
 PUBMED
 REFERENCE 2 (bases 1 to 83906)
 Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
 Direct Submission
 Submitted (08-JAN-1999) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research, 1532-3, Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
 Tel: 81-438-52-3935, Fax: 81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=MLN21
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Grail-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE

```

(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/etRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MLB3 and the 3' clone is MOA2.

FEATURES
     source
       1..83906
         /organism="Arabidopsis thaliana"
         /mol_type="genomic DNA"
         /strain="Columbia"
         /db_xref="taxon:3702"
         /chromosome="3"
         /clone="MLN21"
         /clone_lib="Mitsui P1"
         /product="GDS1-motif lipase/hydrolase-like protein"
         /note="CDS is reported in Acc# AP000416"
         gene_id:MLE3.2"
         /number=1
         /evidence=not_experimental
         complement(join(1714..2552,2810..3098))
         /note="gene_id:MLN21.1"
         /codon_start=1
         /evidence=not_experimental
         /product="transcription factor EREBP-like protein"
         /protein_id="BAB01029.1"
         /db_xref="GI:9279571"
         /translation="MCGGAIISDFIPPRSLRVNTEFIWPDLLKNVKASKRSNKRSD
         FPDLDDEADFOGDDSAFCEDDDYFVNKPEVFTATTKPVASFYVTSYGAYA
         KTVBAEQAEKSKRKNQKRGIRQKRWGAAIRDRPKRSGWLTGTTABEAA
         RYDAARLRIRGKAKVNPPEKNPSVQKPSAKTNLQKSAKPNKSVTLVQOPT
         HLAQQCNRSFGPDGMFMEKPKQMYNNQGLTNSFDAGNNGYQYFSSDQGSNF
         DCFEWSGHGPKTPISSMLVNNNEASFVEETNAAKLKPNSSDDDLMAVLDNALW
         DTPLEVMEAGADAGATQBEENPVELMSLDEINFMLEGDF"
         complement(5735..8062)
         /note="gene_id:MLN21.2"
         /codon_start=1
         /evidence=not_experimental
         /product="subtilisin proteinase-like protein"
         /protein_id="BAB01030.1"
         /db_xref="GI:9279572"
         /translation="MAFFYFFLLTLSSPSSASSNSLTVIVHVDHEAKPSIFPTH
         HWTSPSLASLTSPSIHTYDVTVFHSARLTSQDASQLLDHSHVLSVPEQVRLH
         HTYRPEFLGRLTQAGLLESDFDGSLVIGDVTGWPERPFDRLGLGVPATKWK
         GQICASCDPESACNKLVGARFCGGYEATNGKNETTEFRSPRSDGHTHTASIS
         AGRYVPASTLVGAHGAAPKARLAKYKVCWSGCCYSDILAPATVADGVDVI
         SLVGVVVPIYDIALAIFAGDAIDRGIFFVSASAGNGPALTITNAPWMTTVGAGT
         IDRFANVLGNKMIIGSVYVGGPLDPGRMYPLVYGSLGDEGSSSLCEGLS
         DPNLVKGKIVLCDRGINSRATKEIVRKNGLGMIANGVFDGGLVADCHVLPATSV
         GASGDEIRRYIESKSRSSKHTATIVFKGRLGIRPAPVAVGSFARGPNPTEPI
         LKPDVTPAGNIIAAMPDRIGPSGVTSDNRTEFNILSCTSNACHVSLAALLAAH
         PDNSPAIRALSALTITAYVDNNGEPMDSTGNISVMDYGSQHVHPKAMPDGLVID
         ITSXYINFLLNSGNTRTNIVTITRQADCDGARAGHVGYNLNPFSFVYFOYGESK
         MSTHIRTYNQSDSVSEIKIRPRGTTVTVEPEKLSFRVVGOKLSFVVRVKTTEV
         KLSFGATVETGHIWSDGGRNVTSLVLTLOQL"
         complement(join(10008..10128,10218..11056))
         /note="emb|CAB67676.1"
         gene_id:MLN21.3
         similar to unknown protein"
         /codon_start=1
         /evidence=not_experimental
         /protein_id="BAB01031.1"
         /db_xref="GI:9279573"
         /translation="MGSSSSSSSTKVDTMLDVTYLSALFYDIEFFPIADESYATEI
         HLOEAMFSFGATGVGNHHPVHRNLVTLVQEPKIEAENPEMPSRLCMCMDEK
         PSSDIIFRGNTCTHACTDCTVRYATKIKENASRIKCPDVECTRIEPTCRDLIPK
         DVPDRKEKILCESLISWDKFCYCPKDCSAMVNNNGANDANTVQTECSCHRLFVQCQ
         KVTWHAGICDDEQRGNTKKSSDEDDALLIQMAKQKWRRCFSCCKFYDVKVEGQCH
         IKRISVLWLWICVFERGMSNYSILQSGMYSINLFKRYVF"
         complement(join(12084..12461,12563..12877))
         /note="gb|IAAF02862.1"
         gene_id:MLN21.4
         similar to unknown protein"

```

```

/codon_start=1
/evidence=not_experimental
/protein_id="BAB01032.1"
/db_xref="GI:9279574"
/translation="MQLYRRPFETKTQYRICPAETMVKLIHPDLTTSAGCEATSSPYL
TTEOSFTIWMKSLVFNCTGTVDSKGNIIYRVDNVNSCKREYLMYDLSGHVLFLL
RQKFGLEKTEGWYRSSSATVETTKLEYFRVKNVQVQPNKQSSSSYVNASCRND
EQCYKRWITRGSSLAIEDNCKGLAEVKRQKQSLNGKLDGDDVITMVMVESQVDHSFIIG
LVLAHSLINCIL"
complement(16501..16918)
/note="gb|AAB63539.1"
gene_id:MLN21.5
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
join(18722..19329,19578..20327,20473..20626,20740..21084,
21193..22037,22134..22465,22571..22694,22775..22903,
23041..24525,24603..24869,24957..25283)
/note="emb|CAB36798.1"
gene_id:MLN21.6
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01033.1"
/db_xref="GI:9279575"
/translation="MGTRDSNNRTFSEIVGLIKSWLPWRGEPATVSRDFWMPDQSCRV
CYECDQFTLINRRHCHRGVFCGCKTANSIPFAPSDLTREDWELRVNVCYCFR
QWEGDGGPVSNITELSTSPSTRSSLSSTSTANSSSFALSGMFLGILGNORVHG
DVSLLHSGNEMTSVTKQKRETSRSSFIATDVEDPFRFALNSIRSDDEYDEYAYOT
GSEQFKSHSGDRDECEASPADISDDQVVEPVDFENGLLWVPPEDEBERESA
LFDENVSDEGEGESKLEIITSLWEAANLLKEDMSKSGMDPGGVYKVKCLASGFR
HDSNVKGVCKKNVNRMSSTIEKARLLILGGLEYQVRSNQLSDFILLQOEKDH
PKMAVAKIHAERNILLVEKSVSRFAQEYLLAKDISILVNLKRPLLDRIARCTAQII
LPSVDLSSQKLYGCEVFRVDRPYPEEHGSTQGVKVKVTLMYFHECPKPLGFTLLRG
ANEDELKKVHVQYVQYAAHYLALETSLADEGASPELPLNGSPITVALPKSTSIER
SISTVPGFTVSTVEKSPTMLSCAEPQANSVPVSELSLTNTLSIQNDIPIPYGSGW
QAREINSPFVRSHNLSLNPDRVIESRNSDLSRSVPVTPADKSNPIVADETTNN
SLHSGGQVVKSGQISIMVENQDNGSELTIAQQNNKRPKFTQCKEFPFPPSPD
HQSLTSLRSRSMVKGTVCERSHLRIKYYSFDPKPLGRFLRDLHLPQSYCRSCEMP
SEAHVCHTQTRGSLTISVKQLDYLPLGKEGKIMWHRCLCPRLNGPPATLRVY
MSDAWGLSTGKFLLELSFSNAAASVACCGSHLRDCLAFYCGNAPCFRATIDV
HSVYLPESILSFNVENODTQRETDVEIEAEALLFSEVLANAISQAEKRRRRIEILE
EVLOKAEFEENMQKILHREVNEGQPLVDILELYRIHQQLPQSYMWHRLINASTIL
HKENSDDTKEBNEKPLAKSOTLPENNAGTNSLUGSEVNLNPDGSDGTDGSLNN
VQAEADTNSLYOEKDDGGEVSPSKLPDTSYPLENKVDYRTQSDQIYMKNLSATL
DAWAGIRQTSVETPTNNKVSLLPPTMSNSTPPISEGLMPIDLPEQNEFKVAYPV
SPALPSKNVSESDSVSLVPPFLNFVRSINKNFLSSOKLDTFGEHSPIVSSFREA
ELOGGPRLLLPVGNDIVVPYDDEPISMIAALMSPEYQRTSAEGESLVSYSELN
IPRPVDTTIDFPRSNGSVDESILSSSSSTSLDDPLSYTKALHARVSYGEGDTLQK
VKYTVTCYAKRFEALRGICLSELEYIRLSRCKWGAQGGKSNVFATLDDRTFI
KQVTKTELESFIRPAPAYFKLSEISTKPTCLAKILGIYQVATKQKSGKTKMDV
LIMENLFGTIVKRLVDLKGSRARYNPDSGSKNVLDDQNLTEAMPTSPIFVGNKAK
RILERAVNDTAFIAGDVMDYSILGVDEKNELVIGIIDFLRQYTWDLHESWVKEF
TGILGGRNEAPTVISPKQYKRRFRKAMTYTFLMVPQWSPFNVAANNSKSDQPEETS
QAGTQAE"
complement(join(25752..25887,25980..26104,26220..26384))
/note="gene_id:MLN21.7"
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01034.1"
/db_xref="GI:9279576"
/translation="MAYSTKLYLPLKTPCFVPKTVRAKTLMLQSQSVHVVKVATVSO
POPRTYKTIIVREDPIQGIICVTDNGDEVICGYDGGPCPTQSPMWASYSREVEL
DLLOSQKESLSDNKDQRFIIASEQEFMMKXSSDFEL"
join(28462..28475,28608..28689,28882..28962,29255..29368,
29463..29564,29876..29934,30031..30133,30228..30293,
30387..30479)
/note="gene_id:MLN21.8"
/codon_start=1

```

```

/evidence=not_experimental
/product="20S proteasome subunit PAE-like protein"
/protein_id="BAB01035.1"
/db_xref="GI:9279577"
/translation="MLFRTEDYDRGVNTSPESGRIFQVEYALAIKLGSTAIGVKCKE
GVLVAEKRTISPLIPSSVKEIMEDDHIGAMSLIADARTLVEHARVETQNHPS
GQEWPTVESTIQALCDLARFGKEEESMSRFFGVSLLIAGHDENGSLYTPDSGTF
WQCNAKAGSGSGAUSQSQEQFNKDITIQEAFLAVSLIKQVMEEKYTPNNVDIAKV
APAYHLTYPOEVZAVTSRL"
CDS
complement [join(31003..31685,31773..33996)]
/note="gene_id:MLN21.9"
/codon_start=1
/evidence=not_experimental
/product="pectinesterase-like protein"

Alignment Scores:
Pred. No.: 2,646-39 Length: 83906
Score: 1700.00 Matches: 356
Percent Similarity: 61.95% Conservativeness: 134
Best Local Similarity: 45.01% Mismatches: 257
Query Match: 42.31% Indels: 44
DB: 8 Gaps: 15

```

US-09-806-767-2 (1-775) X AB022220 (1-83906)

QY		6	PhePheLeuCysIlellePheLeuLeuPheCysSerSerSerGluValleuGlnLys	25
Db		8056	TTCCTTCTACTATTTTCTCCCTTGACCTTTCTTCCTCTCTCTCTGCCTCTCTCT	7997
QY		26	Gln-----ThrTyrlleValGlnLeuHisProAsnSerGluThralalysThrPhe	42
Db		7996	TCAAACTCCTTTGACTTACATCGTCAGTCGACCAGCAAGCTTAACCCCTCA---ATCTTC	7940
QY		43	AlaserLysPheAspTtpHisLeuSerPheLeuGlnGluAlaValleuGlyValGluGlu	62
Db		7939	CCCACCTCATTTTCTACTGGTAGACTTCTCTCTTGCCTCACTCAC--	7895
QY		63	GluGluGluGluProSerSerArgLeuLeuTySerTyrglySerAlarIeGluGlyPhe	82
Db		7894	-----TCTTCTCTCCTCTCC---ATTATCCACATACGACACCGGTTTCCATGGCTTC	7844
QY		83	AlaAlaGlnLeuThrGluSerGluAlaGluilleLeuArgTySerProgluValValala	102
Db		7843	TCCGCTAGACTCACTTCACAAGACGCTAGTCAACTCTTAGACCATCCTCATGCTATCTCT	7784
QY		103	ValArgProAspHisValLeuGlnValGlnThrThrTySerTyryLysPheLeuGluLeu	122
Db		7793	GTTATCCCGGAACAGTCCGTCACCTTCGCACACCACCTCGTTCCTCCGTGTTCTTGGTCTT	7724
QY		123	AspGlyPheGlyAsnSerGlyValTrpSerLyseSerArgPheGlyGlnGlyThrIleile	142
Db		7723	AGGTCCACCGACAAGCGGGTCTACTCGAAGAGCTGATTTCCGGTCCGATCTAGTTATC	7664
QY		143	GlyValleuAspThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetPro	162
Db		7663	GGAGTTATCGGATACTGGTGTTCGGCCGAAAGACCTAGCTTTGATGACCGTGGTCTGGT	7604
QY		163	SerIleProArgLysTrpLysGlyIleCysGlnGluGlyCyluserPhaseserSerSerSer	182
Db		7603	CCTGTTCCCATTAATGGAAGGCCCAATGTATCGTTCCCAAGATTTTTCCGGAGTCTGCT	7544
QY		183	CysAsenArgLysLeuilleGlyAlaArgPhePheIleArgGlyHiAargValalaasnser	202
Db		7543	TGTAACCGGTAAACTGTCGGAGCTAGATTCCTTCGGGTGGGTATGAAGAACCAACGCGG	7484
QY		203	ProGluGluSerProAsnMetProArgGluTyIleSerAlaaArgaspSerThrGlyHis	222
Db		7483	-----AAAATGAATGAAACGACTGAGTTTCGCTCTCCGCGTGACTCCGATGGACAT	7433
QY		223	GlyThrHisThrAlaSerThrValGlyGlySerValSerMetAlaasnValleuGly	242
Db		7432	GGGACTCACAGCTTCAATCTCCGCGCGCGCTAGCTGTTTTTTCCGGCTCAACTCTCGGC	7373

Qy	243	AsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValLysVal	262
Db	7372	TAGCCTCACGGCGTGTCTGCTGGATGGCTCCAAAAGCTAGACTAGCCGCTTCAAAGATC	7313
Qy	263	CysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaLalIeAspValAlaIleGln	282
Db	7312	TGTTGGAACTCCGGTGTGTTACGACTCGATATCTTAGCCGCTTTCGACACCGCGGTGCC	7253
Qy	283	AspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAsp	302
Db	7252	GACGGTGTGCATGTCTCTCTCGTTGGAGGCGTGTGTTCTTATTACCTAGAC	7193
Qy	303	ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAla	322
Db	7192	GCTATCGCTATTGGAGCTTTTGGAGCTATTTCAGAGGCATATTTCGCTCTGCTCCGCC	7133
Qy	323	GlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIle	342
Db	7132	GGTNACGGAGGTCCGGTGTCTTAACGGTGACAAATGTTGCTTCGTGGATGACACAGCTC	7073
Qy	343	GlyAlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeu	362
Db	7072	GGAGCTGGAACAATCGATAGGGATTTCGCGCAATGTGAAACTCGTAACTACGGGAAGATC	7013
Qy	363	LeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGlu	382
Db	7012	ATTTCCTGGTGTAGTGTATACGGTGGACCGGGTCTG---GATCGGGTCCAATGTATCCCG	6956
Qy	383	ValIleTyr-----ValThrGlyGlyAsp---LysGlySerGluPheCysLeuArg	398
Db	6955	CTTGTTTACGGTGGTAGTGTATTACTAGGGCGGATGGTGTACTCATCGTCTCTGTGTCCTTGA	6896
Qy	399	GlySerLeuProArgGluGluIleAlaGlyLysMetValIleCysAspArgGlyValAsn	418
Db	6895	GGCTCGTGGATCCGAATTTAGTCAAGGGAAGATCGTCTCTTGTGATAGAGGAATCAAT	6836
Qy	419	GlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAla	438
Db	6835	TCTAGACCAACCAAGGTGAGATCGTACCGCAAAAATGGAGCTTGGGGATGATTATAGCG	6776
Qy	439	AsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeu	458
Db	6775	AATGGTGTGTTGACGGTGAAGGTTTAGTAGCTGATTGCCACGCTGTACCGGACATCT	6716
Qy	459	IleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLys----	476
Db	6715	GTTGGTCTCTCGGAGAGATGAGATTCGTAGGTATATCTCTGNAATCATCCAAATCTCGT	6656
Qy	477	-----ProTyrAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArg	492
Db	6655	TCATCGAAACATCCACCGGTACGATTGTTTCAAAGGTACCCGACTGGGATCCGACCT	6596
Qy	493	AlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeu	512
Db	6595	GCACCGGTGTTCGATCTTTCTCTGCTCGTGGTCTTAATCCAGAGACGCCGCGAGATTCCT	6536
Qy	513	LysProAspVetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGly	532
Db	6535	AAACCGGATGTATCGCACTCTGTTTGAATATTCTAGTCTGCTTGGCTGACCGGATTTGGT	6476
Qy	533	ProThrGlyLeuProTyrAspSerArgValAsnPheThrValMetSerGlyThrSer	552
Db	6475	CCATCTGTTGTTACTTCTGATAACCGGAGAACTGAGTTCAAACATTTTATCAGGCACCTCG	6416
Qy	553	MetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrp	572
Db	6415	ATGGCGTCCCGCACGCTGTCTGTCTAGCTGCTTTGCTTAAGCGGGCTCATCCGGATTTGG	6356
Qy	573	SerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGly	592
Db	6355	AGTCCGCGCGGATAGATCAGCATGATAACCAACCGGCTTACACGGTTGATATAACACGGGT	6296
Qy	593	LysAlaIleLysAsp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAla	609

/notes="synonym: T6P5.4; identical to GB:AF043519;
supported by cDNA: gi.15450622.gb.AV052679.1"
join(18080..18214,18297..18393,18473..18554,18619..18715,
18791..18849,18932..19014,19321..19416,19606..19665,
19874..20180)
/gene="At2g05840"
join(18139..18214,18297..18393,18473..18554,18619..18715,
18791..18849,18932..19014,19321..19416,19606..19665,
19874..19966)
/gene="At2g05840"
/codon_start=1
/product="20S proteasome alpha subunit A2 (PAA2)"
/protein_id="AAC95161.1"
/db_xref="GI:4006820"
/translation="MSRGGAGYDRHITIFSEGRFLQVEYAFKAVKAGLTSIGVRG
KQSCVTVKQKVPDKLLDQSSHLFPVTKYLGILGATGNTADRSLSVQARNEAEFR
FOYQEMPAIDAKITADSKQVYTHAYMRPLGVAMVILGIDDEERGPLYKCDPAGHF
YGHKATSAAGKQEAQVNFLEKMKENPAFTYDTVTQTALQSALQVLDKFRATEIEVG
VWADDPFLRSRLTEIDELHTAISERD"
complement(23222..23330)
/rpt_family="AT_rich"
complement(23980..24022)
/rpt_family="AT_rich"
complement(24550..24574)
/rpt_family="AT_rich"
complement(24765..24786)
/rpt_family="AT_rich"
complement(24944..25022)
/rpt_family="AT_rich"
25297..25317
/rpt_family="AT_rich"
25827..25856
/rpt_family="AT_rich"
complement(25903..25966)
/rpt_family="AT_rich"
complement(26687..28373)
/gene="At2g05850"
/notes="synonym: T6P5.5"
complement(join(<26687..27696,27802..27909,28028..>28373))
/gene="At2g05850"
complement(join(26687..27696,27802..27909,28028..28373))
/gene="At2g05850"
/codon_start=1
/product="putative serine carboxypeptidase II"
/protein_id="AAC95162.1"
/db_xref="GI:4006820"
/translation="MGKQDWSVTACIFLSLSIASOIHCSOTHPSPHKGAGLGGDT
SHNSYSRNLVSLKEDLIEKLPQSGISPROYGVVAVNEPATRFLYYVFEAIK
PSKSTLVLFNFGPCSSVGFAPFELGPRVHSDCKLYLRNPISWNNEANMLFPEG
PISVGSYSTPDDWEIFGEQADKLTAEDNPMPLVNWLERFEYKGRDVIYISGQSYAG
HYTPQAQILHNNQTFINLRGISIGNFLDLILLADNENKFIILSHGLVSQKQFEY
SKYCDPANDMDPCPKIPKFSIEHNKHLVDVNIYAPVCLNSTLSPEPKCTTMEVD
PCRSNVYKALNSGVQEAHANTTKLPYEWKACHYLSNVWDADKQASMPILHDL
MGSGRVLYLSGVDAALPFTATMAVLTKMLTVNWRPFWFTGGQLGFTEDYENL
TYARVSGSHSVLPDQFVALNLFTSFIRNPLPQP"
28927..28981
/rpt_family="AT_rich"
complement(29152..29229)
/rpt_family="AT_rich"
complement(31013..32021)
/gene="At2g05860"
/notes="synonym: T6P5.6"
complement(join(<31013..31223,31303..31405,31542..31806,
31884..>32021))
/gene="At2g05860"
complement(join(31013..31223,31303..31405,31542..31806,
31884..32021))
/gene="At2g05860"
/codon_start=1
/product="putative replication protein A1"
/protein_id="AAC95163.1"
/db_xref="GI:4006821"
/translation="MESFHLKSLKSRGWCIRGRVWRTFLVSLVPSKVMGLILAD

Alignment Scores:

Pred. No.:	6.34e-99	Length:	121160
Score:	1697.50	Matches:	370
Percent Similarity:	60.47%	Conservative:	115
Best Local Similarity:	46.13%	Mismatches:	241
Query Match:	42.25%	Indels:	76
DB:	8	Gaps:	16

US-09-806-767-2 (1-775) x AC005970 (1-121160)

Qy	11	IlePheLeuLeuPheCysSerSerSerGluLeuLeuGlnLysGlnThrThrIleVal	30
Db	56069	CTCTCTCTCTCTCTCCACACACCGCC-----AAAAAACCTCATCATC	56025
Qy	31	GlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTTPHisLeu	50
Db	56024	CGCGTCAAT---CACTCCGATAAACCGGAATCATCTCTACTCACACGATTCGTACACA	55968
Qy	51	SerPheLeuGlnGluAlaValLeuGlyValIgluGluGluGluGluProSerSerArg	70
Db	55967	TCTCAACTCAATCA-----GAATCATCT	55944
Qy	71	LeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGlu	90
Db	55943	CTTCTCTACACTTACACACACTCTCTCCATGGCTTCTCCGCTTACCTCGACTCCGCGAA	55884
Qy	91	AlaGlu---IleLeuArgTyrSerProGluValValAlaValArgProAspHisValLeu	109
Db	55883	GCCGATTCCTCTCTCTCTCTCAAACTCAATCTCGATATCTTCGAAGATCTCTCTAC	55824
Qy	110	GlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGly---PheGlyAsnSer	128
Db	55823	ACATTCACACTACGCGTACTCTCGAGTTCTCGGCTCAATCCGAAATTCGGTGTTTAC	55764
Qy	129	GlyValTyrSerTyrSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGly	148
Db	55763	GATCTCGGTTCTTCTTCT-----AACGGCGTTATCATCGAGTTTATAGATACTGCG	55713
Qy	149	ValTyrProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTyr	168
Db	55712	GTTTGGCGCTGAATAGAACCTCGATGATGATGATGATGATGATGATGATGATGATGATG	55653
Qy	169	LysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLysLeuIle	188
Db	55652	AAAGAGAAATGTGAATCTGGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC	55593
Qy	189	GlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsn	208
Db	55592	GGACTAGAGCTTCTCCAAAGGATTTCAATGGCTTCTGGTGGTGGTGGTGGTGGTGGTGG	55533
Qy	209	MetProArgLysThrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer	228
Db	55532	-----CGTCAATCTGTTTCTCTCTCGTATGTTTCACGACATGGAACACATCTTCACT	55479
Qy	229	ThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAla	248
Db	55478	ACCCCGCGGATCCGCGTGTAGAACGCTATCTCTCGTTCCTCGGTTACGCGCGGTACGCG	55419
Qy	249	ArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTyrPheAsnGlyCys	268
Db	55418	AGAGGTATGGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	55359
Qy	269	TyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeu	288
Db	55358	TTTGGATCTGATATCTAGCTGCTATGATCGAGCTATCTTCTGATGCTGCTGCTGCTGCT	55299
Qy	289	SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThr	308

Db 55298 TCGTTATCTCTGTGTGTCTGCTCCGTAATATTCGCGATACGATTCGATTCGATTCGAGCG 55239
 Qy 309 PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle 328
 Db 55238 TTTTCGGCTATGGAGAGAGGTGTTTGTCTCTGTTCTCTGTAATAGTGTCTACT 55179
 Qy 329 GluSerSerValAlaAsnThrAlaProTTPValSerThrIleGlyAlaGlyThrLeuAsp 348
 Db 55178 AGACCTCTCTGTGCAATGTTGCTCTGCTGTTGATGACTGTTGCTGCTGTTAGAT 55119
 Qy 349 ArgArgPheProAlaValAlaValArgLeuAlaAsnGlyIleLeuLeuTyrglyGluSerLeu 368
 Db 55118 AGACATTTTCGGGTTTTTCGCAATCTCGTAACGGGAACGACTTACCGGTGTTTCGCTG 55059
 Qy 369 TyrProGlyIleGlyIleLysAsnAlaGlyArgGluValGluValIleTyrglyValThrGly 388
 Db 55058 TATAGCGGTAGGAATG-----GGACGAAGCCGCTTGAATGTTTATATAAAGGG 55005
 Qy 389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly 408
 Db 55004 AATAGTAGTTCGAGTAATCTTTGTTTACCTGGTTGCTTCGATTTCGAGTATGTTGCTGG 54945
 Qy 409 LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys 428
 Db 54944 AAGATTCTGTTGATAGAGGTGTTAATGCTAGAGTTGAGAAGAGAGCTGTGTTAG 54885
 Qy 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrCluIleAsnGlnGluAspSer 448
 Db 54884 GATCTGCTGTTAGGAGTGAATGCGGAATACGCTCGGAGTGAGAGGAGCTGTG 54825
 Qy 449 IleAspValHisLeuLeuProAlaThrLeuIleGlyTyrrThrGluSerValLeuLys 468
 Db 54824 CGCGATAGTCAATTGCTTCCGCGATCGCTGTAGGAGAGAGACTGGTATTTACTTAG 54765
 Qy 469 AlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyThrValIle 488
 Db 54764 GAGTATGTTAAGTCAGATCTAAACCAACCGCTCTCTGTTTAAAGGAACGGTCTT 54705
 Qy 489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508
 Db 54704 GACGTTAAGCGCTCTCTGTTGCTGTTTGTAGCTCGAGAGGCTCTAATCTGTTACT 54645
 Qy 509 ProSerIleLeuLysProAspMetIleAlaProGlyValAlaSerIleAlaIleAlaTTPPro 528
 Db 54644 CCGAATCTTGAAGCCTGATGTTATTTGCTCGAGGTTAATATTTTGGCTGCTGCT 54585
 Qy 529 GlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMet 548
 Db 54584 GACGCTATGCTCTACTGTTCTTGACAAGGACTCTAGAGGACTCAGTTCAACATCAG 54525
 Qy 549 Ser----- 549
 Db 54524 TCAGGTACTCAACTTGAAACTTGATTTGATTACATATTTTGTATTGATTATATCTAT 54465
 Qy 550 -----Glyth 551
 Db 54464 GATTGTAGAGAGGGTTTGTGTTTGTGACATGAAAGTCTCGATTATTTGATTAGTAC 54405
 Qy 551 rSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAs 571
 Db 54404 GTCATGTATGCCACACATCAGTGTGTTAGCGGGTCTTTTGAAGAGCAGCTCACCCTGA 54345
 Qy 571 nTTPSerProAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGly 591
 Db 54344 GTGGAGTCCGAGTCTATCAATCAGCTCTCATGACTACAGCTTACGTTCTTGACAACAC 54285
 Qy 591 nGlyIleAlaIleLysAspGly-----AsnLysProAlaGlyValPheAlaIleGlyAl 609
 Db 54284 CAACGCTCTCTCCATGATGTGTCAGACAACAGCCTATCTAACCATATCTCAGCGCT 54225
 Qy 609 aGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProVa 629
 Db 54224 GGGCCATGTAGATCCCCAAAAGGCTCTCTCACCAGGCTTGTCTAGACATCTCAACCGA 54165

Qy 629 lAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleTh 649
 Db 54164 GGAATACATCAGGTTTTTGTCTCTAGACTACACAGTCGATCAGATTTGTCGATTGT 54105
 Qy 649 rHisLys---AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAs 668
 Db 54104 GAAGCGACTAGCGTTTAACTCTCGAAGAAGTTC---TCAGATCTGCT---CAGCTCAA 54051
 Qy 668 nTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgAr 688
 Db 54050 CTACCCAGTTTTCGGTTTTTGT---GGGGGTAAGAAGAGTTGTCGGTACATCTCGGA 53994
 Qy 688 gValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluG1 708
 Db 53993 AGTAACAATCTGTCGACGACGCTCGTTTACAAAGTGACGGTTAATGGAGCTCCTAG 53934
 Qy 708 yIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSe 728
 Db 53933 TGTGGAAATCTCTGTTAAACCATCGAACTTTCGTTTAAAGCGTGGGAGAGAAGAG 53874
 Qy 728 rTyrArgValTTPheValLeuLysLysLysAsnArgGlyGlyLysValAlaIleSerPheAl 748
 Db 53873 GTACACAGTCACTGTTGTAGCAAGAAAGGAGTGAGTATGACGAACAGGCTGAGTTT-- 53816
 Qy 748 aGlnGlyGlnLeuThrTTPValAsnSerHisAsnLeuMetGlnArgValArgSerProI1 768
 Db 53815 ---GGTTCATCACTTGGACCAATCCGACAC-----GAAGTGAGAGTCCCGT 53769
 Qy 768 eSer 769
 Db 53768 TGCA 53765

RESULT 10
 AY084387
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2638)
 Haas,B.J., Volfovsky,N., Town,C.D., Troughan,M., Alexandrov,N.,
 Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
 Full-length messenger RNA sequences greatly improve genome
 annotation
 Genome Biol. 3 (6), RESEARCH0029 (2002)
 22088475
 12093376
 2 (bases 1 to 2638)
 Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
 Feldmann,K.
 Full-length cDNA from Arabidopsis thaliana
 Unpublished
 3 (bases 1 to 2638)
 Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
 Feldmann,K.
 Direct Submission
 Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
 Malibu, CA 90265, USA
 This clone sequence is one of 5,000 Ceres full-length cDNAs made
 available to TIGR and Genbank. The following quality assessment of
 this set was done by comparison with known proteins: two percent of
 the clones are estimated to be 5'-truncated; less than one percent
 are 3'-truncated; approximately two percent represent alternative
 splice variants, including unspliced introns and spliced exons; one
 percent may contain premature stop codons; five percent may have
 frame shifts in a coding region. A sequence is considered to be
 5'-truncated if it lacks the translation initiation start (ATG). A


```

Db 1621 GCACCGGTTGTTGCACTTTCTCTGCACGTGCTTAATCCAGACGCGGAGATTCTT 1680
Qy 513 LysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGly 532
Db 1681 AACCCGATGTAATCCACTGTTTGAATATCTAGCTGCTTGGCCTGACCGGATTGGT 1740
Qy 533 ProThrGlyLeuProTyrAspSerArgValAsnPheThrValMetSerGlyThrSer 552
Db 1741 CCATCTGGTGTACTTCTGATAACCGGAGAACTGAGTTCAACATTTTATCAGGCACCTCG 1800
Qy 553 MetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrp 572
Db 1801 ATGGCGTGGCCCATGTTTCTGTCTAGCTGCTTGTCTTAAGCGGCTCATCCGATTGG 1860
Qy 573 SerProAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGly 592
Db 1861 AGCCCGCGCGGATAAGATCAGCATTTGATGACACGGCTTACACGGTTGATAACACGGCT 1920
Qy 593 LysAlaIleLysAsp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAla 609
Db 1921 GAGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1977
Qy 610 GlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProVal 629
Db 1978 GGTGATGTTTCAACCAACAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2037
Qy 630 AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThr 649
Db 2038 GATTATATCAACTTGTGTGTTTAACTTCACTACATGGAACCAACATTTGTGACGATAACC 2097
Qy 650 HisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyr 669
Db 2098 GTGCGCAAGCGGACTGCGAGCGGTGCGAGCGGCGGACACGTCGGAATTTGAACATC 2157
Qy 670 ProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThr----- 686
Db 2158 CGTCGTTCTGCTGCTGTTTTCAGCAGTATGAGAGAGTAATAATGTCGACATTTTCATT 2217
Qy 687 ArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaPro 706
Db 2218 AGGACTGTGACAAATAGAGCATTCGGATTCGGTTTACGAGATTAAGATTAGGCGGCT 2277
Qy 707 GluGlyLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThr 726
Db 2278 AGAGGACTACATGACGTTGAGCGGAGAGCTATCCTTCAGACGGTGGGCGAGAA 2337
Qy 727 LeuSerTyrArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAlaSer 746
Db 2338 CTGAGT-----TTTGTGTAGGTTAAGACACAGAGGTGAGTTGTACCA 2385
Qy 747 PheAla-----GlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGln 762
Db 2386 GGAGCGACAAATGTGAGACTGCTCATATGTTTGTGTCAGATGG-----AAACGG 2436
Qy 763 ArgValArgSerProLysSerValThrLeuLys 773
Db 2437 AATGTGACAAAGTCCCTTGGTGGTCACTCTTCAA 2469

RESULT 11
LESBT2
LOCUS 2650 bp mRNA linear PLN 18-FEB-2000
DEFINITION L.esculentum mRNA for serine protease, SBT2.
ACCESSION X98930
VERSION X98930.1 GI:1771161
KEYWORDS SBT2 gene; serine protease.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 Janzik, I., Macheroux, P., Amrhein, N. and Schaller, A.

```

```

LeSBT1, a subtilase from tomato plants. Overexpression in insect
cells, purification, and characterization
J. Biol. Chem. 275 (7), 5193-5199 (2000)
MEDLINE 20138270
PUBMED 10671566
REFERENCE 2 (bases 1 to 2650)
AUTHORS Schaller/A.
DIRECT SUBMISSION
SUBMITTED (01-JUL-1996) A. Schaller, Federal Institute of
Technology, Institute of Plant Sciences, ETHZ, LFW E51,
Universitaetsstrasse 2, CH-8092 Zuerich, SWITZERLAND
FEATURES
Location/Qualifiers
1..2650
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/sub_species="UC82b"
/db_xref="taxon:4081"
/clone_lib="Lamba ZAP"
/dev_stage="leaf"
1..2650
/genes="SBT2"
115..2443
/genes="SBT2"
/function="serine protease"
/codon_start=1
/evidence=not experimental
/protein_id="CAA67430.1"
/db_xref="GI:1771162"
/db_xref="SPTREMBL:P93205"
/translation="MAGMLKCMFFVSVCLAINLAKSPNTKKTIIQMDKWAKPDV
FVHVOMVSLKSVLPSVTEVTGDEBERILYSYQATPHGVAQLSEEVKKLOER
NGVLAYPEIKYOLHTTRPFLGLDREDSSKLWADRLSDHNVIGVLDTGIPESPS
FNDGWTSPVSHMKVCETGRGPEKHCCKIIVARVFRGVAAAGSKINERGFKSA
RDQGHGTHGTAGVSVRGANLLGAYATGAKMAPARVAAKYKVCWVGCCRSSDIL
SAVDQAVADGVNLSISGGVSSVNRSLIAAFAGMAGKEVFSVCSAGNGGDDPISL
TNYSPWITTVGASTMDRDPATVLTGKIVTGSALYKGMNLSLTKQYPLIYLGNS
SNLPSLLCLDGTLDKASVAGKIVICDRGISPRVQGVVKEAGVGMILNTAANGE
ELVADSHLPAVAVGREGRAIKLYAAGRSATATLFLGTGKIGRSPVVAFFSRGP
NFSLLEIKEDMVAQVNIILAGTNGALGPSSLIDQRRTNFNLISGTSMSCHVSGIA
ALAKARHPDSPAALKALMTAYVDNTIKSLKQASVTFPIYDHGAGHVNPKRAV
DGLIYDIGAQDYFEFLCTQELSPQVMVFGEKSNVTHSLANPDGLNYPALSAVFP
EKTLSMLTHTRTVTVGSPISNHYVVSFAKGVAVKVEPERLNTFSKQKLSYKVTF
KTVSRQAKPEFGSLWKDGTGKVRSPAIATWLASV"
BASE COUNT 756 a 511 c 618 g 765 t
ORIGIN
Alignment Scores:
Pred. No.: 5,45e-100 Length: 2650
Score: 1679.00 Matches: 359
Percent Similarity: 61.97% Conservative: 125
Best Local Similarity: 45.97% Mismatches: 261
Query Match: 41.79% Indels: 36
DB: 8 Gaps: 16
US-09-806-767-2 (1-775) x LESBT2 (1-2650)
Qy 6 PhePheLeuCysIle-----IlePheLeuLeuPheCysSerSerSerSerGluIle 22
Db 146 TTCTTTGATCAGTATGCTTGTCTTATCAACTGACCAATATGATAGCCAAACACT----- 199
Qy 23 LeuGlnLysGlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLys----- 40
Db 200 -----AAGAAACTTACATTATCCAAATG-----GATAAATGGGTAAACACGAT 244
Qy 41 ThrPheAlaSerLysPheAspTyrPheLeuSerPheLeuGlnGluAlaValLeuGlyVal 60
Db 245 GTATTCGTCACCATGTTCAATGCTACTCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 304
Qy 61 GluGluGluGluGlu-----GluProSerSerArgLeuLeuTyrSerTyrGlySerAla 78
Db 305 ACAGAGGTAGAAAAACACGGGATGGAGAGAGGATACCTTTTACAGCTACACGACGCTCT 364
Qy 79 IleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluIleLeuArgTyrSerPro 98

```


DEFINITION Lycopersicon esculentum sbt2 gene.
 ACCESSION AJ006379
 VERSION 1 GI:3687306
 KEYWORDS sbt2 gene; serine protease; subtilisin-like protease.
 SOURCE Lycopersicon esculentum (tomato)
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 REFERENCE 1
 AUTHORS Meichtry,J., Amrhein,N. and Schaller,A.
 TITLE Characterization of the subtilase gene family in tomato (Lycopersicon esculentum Mill.)
 JOURNAL Plant Mol. Biol. 39 (4), 749-760 (1999)
 MEDLINE 99277592
 PUBMED 10350089
 REFERENCE 2 (bases 1 to 5197)
 AUTHORS Schaller,A.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAY-1998) Schaller A., Institute of Plant Sciences, Federal Institute of Technology Zurich, Universitätsstr. 2, CH-8092 Zurich, SWITZERLAND
 FEATURES
 source
 1..5197
 /organism="Lycopersicon esculentum"
 /mol_type="genomic DNA"
 /cultivar="VFw8"
 /db_xref="taxon:4081"
 1793..4446
 /gene="sbt2"
 /gene="sbt2"
 /gene="sbt2"
 /evidence="experimental"
 1908..4235
 /gene="sbt2"
 /function="serine protease"
 /codon_start=1
 /product="subtilisin-like protease"
 /protein_id="CAA07000.1"
 /db_xref="GI:3687307"
 /db_xref="SPRMBL:P93205"
 /translations="MAGMLLKMFVSVCLAINLAKSPNTKTVIIQMDKWKPDV
 FVDHVQVSLVKSVPSPTEVEKTDGGERILYSQTAFPHGVAQLSEEVKQLQER
 NGVAFFPEIKQLHTIRSPFLGLEDREDSKLDRLSDHNVIVGLDITGWPSPS
 FNDGMSVPEHWKVCETGREGKSKKIVGARVFRGVEAASGKINERGEFKSA
 RDQDGHGTAGTVAGSVRGANLGYCTARGAPVAAYKVCWVGCGGDPDIL
 SAVDQAVADQVNTLSISLGGVSVNRDLSIAAFGAMKEGVFVSCVAGGPDPSIL
 TNYSPMTTIVGASTMDRDPATVELCTGKIVTGSALYKGRMNLSTQKQYPLIYLSNS
 SNLMPSSCLDGLDKASVAGKVICDRGISPRVQKGVYKAGGVGMILNTANGE
 ELVADSHLLPDAVAGREGRAIKLYAAGRSATATLFLGLKIGIRPSPVVAFFSRGP
 NFLSLRPLDMVAPGVNTLAGWTGALGPSLIPIDQRTNPNLISGTSNCPHVSIA
 ALLKARHPDWSPAIKSALMTAYVHDNTYKSLKDASSVTPSTPYDHGAGHVNPRKAV
 DPLIYDIGAODYFEFLCTCELSPLQMLVFGKFSNRTCHHSANPGDLNYPALSAVFP
 ETKLKLMLTHRTVNVGPISNYHVVSAPFGAVYKVEPERLNTFTSKNQKLSYKTFP
 KTVSRQAPFGLSLWKDGHKVRSPIALTWLASV"
 BASE COUNT 1701 a 864 c 964 g 1668 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,37e-99 Length: 5197
 Score: 1679.00 Matches: 359
 Percent Similarity: 61.97% Conservative: 125
 Best Local Similarity: 45.97% Mismatches: 261
 Query Match: 41.79% Indels: 36
 DB: 8 Gaps: 16

US-09-806-767-2 (1-775) x LES6379 (1-5197)

Qy 6 PhePheLeuCysile-----IlePheLeuLeuPheCysSerSerSerSerGluile 22
 |||||
 Db 1938 TTCTTTGATCAGTATCTTGTATCACTTAGCCAAATGAGCCCAACACT----- 1991
 |||||
 Qy 23 LeuGlnLysGlnThrTrpIleValGlnLeuHisProAsnSerGluThrAlaLys----- 40
 |||||

1992 -----AAGAAACTTACATTATCAAAATG-----GATAAATGGCTAAACACAGAT 2036
 |||||
 Qy 41 ThrPheAlaSerLysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyVal 60
 |||||
 Db 2037 GTATTGTCGACCAATGTTCAATGGTACTCATCGTCTGTAATAATCAGTACCTACCATACACA 2096
 |||||
 Qy 61 GluGluGluGluGlu-----GluProSerSerArgLeuLeuTrpSerTrpGlySerAla 78
 |||||
 Db 2097 ACAGAGGTAGAAAAACAGGGGATGGAGAGAGGATCTTTACAGCTACCACTGCT 2156
 |||||
 Qy 79 IleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluLeuLeuArgTrpSerPro 98
 |||||
 Db 2157 TTTCATGCTGTGTCAGCTCAGCTTAGTGAAGACAGTGAAGAAATTTACAAGACGAAAT 2216
 |||||
 Qy 99 GluValValAlaValArgProAspHisValLeuGlnValGlnThrTrpSerTrpLys 118
 |||||
 Db 2217 GGGGTTTGTAGCTGTTTCTCTGAGATTAAAGTATCAGTTATACCATCAGTAGAAGTCCTCTG 2276
 |||||
 Qy 119 PheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGln 138
 |||||
 Db 2277 TTTCTTGGCTTGACCGTGAGGATAGCAGTAAGTTGTGGCGGATAGATTATCTGACCAT 2336
 |||||
 Qy 139 GlyThrIleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAsp 158
 |||||
 Db 2337 AATGTAATTGGTGGTGTCTTGATCTGGGATTTGGCGGAAAGTCCGAGTTTCAATGAT 2396
 |||||
 Qy 159 ThrGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPhe 178
 |||||
 Db 2397 AGTGAATGACTAGTGTTCATCTCACTGGAAGGGGTTTGTGAACAGGAGGGGTTT 2456
 |||||
 Qy 179 SerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArg 198
 |||||
 Db 2457 GAGAAGCATCATTCGACGACGAAATTTGTTGGTCTAGAGTGTTTTCCTGGGTGATGAA 2516
 |||||
 Qy 199 ValAlaAsnSerProGluGluSerProAsnMetProArgGluTrpIleSerAlaArgAsp 218
 |||||
 Db 2517 GCAGCTTCTGGT-----AAATCAATGAAGGGGTGAGTCAATCGCGAGAGAT 2567
 |||||
 Qy 219 SerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMerAla 238
 |||||
 Db 2568 CAAGATGGCATGCTACTCATCTGGAACAGTCTGAGGTCTGTAGTCCGTGGTGTCT 2627
 |||||
 Qy 239 AsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAla 258
 |||||
 Db 2628 AATCTTTGGGTTATGCTTATGCTAGTACTGCTAGAGAAATGGCCCTGTGGCGAGAGTTGCA 2687
 |||||
 Qy 259 ValTrpLysValCysTrpPheAsnGlyCysTrpSerSerAspIleLeuAlaIleAsp 278
 |||||
 Db 2688 GCATATAGGTGTGTGGGTGGGTGATGTTTGTAGCTCAGATATACTTCTTCGAGTTGAT 2747
 |||||
 Qy 279 ValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIlePro 298
 |||||
 Db 2748 CAAGCTGGCTGATGGGTAAATATCTCTCAATTTCTTGGGTGGGTGGGTTCGCTCT 2807
 |||||
 Qy 299 LeuTrpAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerVal 318
 |||||
 Db 2808 TATAATCGTGACAGTTTGTCAATTCGCTTTTGGGGCCATGGAGAAAGGGGTTTGT 2867
 |||||
 Qy 319 IleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrp 338
 |||||
 Db 2868 TCATGTTTCAGCAGAAATGGTGGCCCTGATCCCAATTAGTTGACAAATGTGTCCACATGG 2927
 |||||
 Qy 339 ValSerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAla 358
 |||||
 Db 2928 ATCACTACTGTAGAGCAAGCACTATGATAGAGATTTTCAGCAACTGTTGACTGGGG 2987
 |||||
 Qy 359 AsnGlyLysLeuLeuTrpGlyGluSerLeuTrpProGlyLysGlyIleLysAsnAlaGly 378
 |||||
 Db 2988 ACGGGAGAGATTGACCGGAGCTTCACTTTATAAGGTAGGATGAATCTTTCACACAA 3047
 |||||
 Qy 379 ArgGluValGluValIleTrpValThrGlyGlyAspLysGly-----SerGluPhe 395
 |||||

Db 3048 AAACAGTATCTCTTTGATTACTTAGGAAGTAATTCGAGCAATCTCATGCCGAGCTCAATTG 3107
 Qy 396 CyeLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArg 415
 Db 3108 TGCTTAGATGGTACCTTAGATAAAGCTTCAGTTGCTGGAAATTTGTGATATGTGCACAGG 3167
 Qy 416 GlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMet 435
 Db 3168 GGAATAGTCTCGATTCMAAAGGGCAAGTTGTAAAGAGAGCTGGAGGAGTAGAATG 3227
 Qy 436 IleLeuAlaAsnThrGluLeuAsnGlnGluLeuAspSerIleAspValHisLeuLeuPro 455
 Db 3228 ATTTTGACAAACACTGCAGCTAATGTGAAGAACTGGTTGCTGATTCTCACCTTCTTCG 3287
 Qy 456 AlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrVal 475
 Db 3288 GCAGTAGCAGTGTGGTGAAGAGAGAGAGCAATCAAGCTCTACGCG---GCAGGTAGA 3344
 Qy 476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
 Db 3345 AGCGTACTCCACTTTGAGATTCTTGGAATCTTGGACTAAATTAGGATCAGGCTTCGCCGTG 3404
 Qy 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
 Db 3405 GTTGTCTGCATTTTCATCCAGAGCACTAATCTCTCAGCTAGAGATTCTTAAGCCAGAT 3464
 Qy 516 MetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGly 535
 Db 3465 ATGGTGCCACAGGAGTGAATATCTTGTGCTGGTGTGACCTGGTGCATTTGGTCCGCTAGT 3524
 Qy 536 LeuProTyrAspSerArgValAsnPheThrValMetSerGlyThrSerMetSerCys 555
 Db 3525 TTCCCAATTGATCAACGAGGAGCAATTTCAATATTCTGTGGAACTTCAATGTCATGT 3584
 Qy 556 ProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAla 575
 Db 3585 CCACATGTTAGTGCATTTGCTGCTTCTCAAGCCAGACATCCAGATTTGGAGTCCAGCA 3644
 Qy 576 AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIle 595
 Db 3645 GCAATTAATCCGCACTCATGACAACTGCTTATGTCATGATAACACTTACAATCTCTA 3704
 Qy 596 LysAspGlyAsnLys-----ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsn 613
 Db 3705 AAAGATGCCCTCCGCTGACTCTCTCGACCCCATATGATCAGCGTCAGACAGCTCAAT 3764
 Qy 614 ProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThr 633
 Db 3765 CCAAGGAAGCGGTGATCCCGTCTGATTATGACATAGGCGGCACAAAGATTACTTCGAG 3824
 Qy 634 TyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIle---ThrHisLysAsn 652
 Db 3825 TTCTTGTGCAGCGCAAGCACTCAGCCCTTCACAGCTAATGTTTTCGGCAAGTTCTCAAC 3884
 Qy 653 ValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIle 672
 Db 3885 AGAATTCGCCACCACTCACTT---GCTAATCCAGGG---GACTTGAACCTACCCAGCCATC 3938
 Qy 673 AlaValIlePheLysArgGlyLysThrThrGluMetIleThr-----ArgArgValThr 690
 Db 3939 TCTGCAGTTTTTCTGAAAAAGACAAACTTTCATATGCTGACCTTCATGAACAGTCACCT 3998
 Qy 691 AsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLys 710
 Db 3999 AATGGGTGTCCTCCCATCTCAATATCATGTTGTAGTCTCAGCATTTAAAGCCGAGTC 4058
 Qy 711 ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg 730
 Db 4059 GTGAAGGTTGAGCCAGAACGATTGAATTTTCCACCAGCAAAACCCAGAACTGCTTTACAAG 4118
 Qy 731 ValTrpPheValLeuLysLysLysAsnArgGlyGlyValAlaLysPheAlaGlnGly 750
 Db 4119 GTCACATTCAAAGCGGTATCTCGTCAA-----AAAGCACCTGAATTT-----GGA 4163

Qy 751 GlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProfileSerVal 770
 Db 4164 TCCCTAATATGGAAGATGGA-----ACACATAAGTAAGAGCCGATTCGAATC 4214
 Qy 771 Thr 771
 Db 4215 ACA 4217

RESULT 13

AY142612

LOCUS

DEFINITION

Arabidopsis thaliana putative subtilisin serine protease ARA12

(At5G67360) mRNA, complete cds.

ACCESSION

AY142612.1

VERSION

AY142612.1

KEYWORDS

FUT CDNA.

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Eukaryota; Viridiplantae;

Spermatophyta; Magnoliophyta;

rosids; eurosids II; Brassicales;

Arabidopsis

REFERENCE

1 (bases 1 to 2305)

AUTHORS

Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Arabidopsis Open Reading Frame (ORF) Clones

Unpublished

2 (bases 1 to 2305)

AUTHORS

Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Direct Submission

Submitted (16-AUG-2002) Plant Gene Expression Center, 800 Buchanan

Street, Albany, CA 94710, USA

COMMENT

The RIKEN Genomic Sciences Center (GSC) members carried out the

collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN

Arabidopsis Full-length cDNA': Seki, M., Narusaka, M., Ishida, J.,

Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,

Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PDEC (SSP) Consortium members constructed and

sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada, K.,

Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M.,

Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C.,

Wu, H.C., Yu, G., Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J.,

Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tripp, M.G., Wu, T.,

Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PDEC) and Seki, M. (RIKEN GSC) contributed equally

to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP

/PDEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis

genome submitted to GenBank.

Location/Qualifiers

1..2305

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="5"

/clone="U09277"

/note="This clone is in pUNI 51.

ecotype: Columbia"

1..2305

gene

```

/gene="At5g67360"
1. .2274
/gene="At5g67360"
/notes="contains subtilase domains PF00082 and a PA domain
PF02225"
/codon_start=1
/evidence=experimental
/product="putative subtilisin serine protease ARA12"
/protein_id="AA13181.1"
/db_xref="GI:23296832"
/translation="MSSSPFUSSTAFFLLCLGFCVSSSSSDOCTYIVHMAKQMPSS
FDLHSNWYDSSLASIDPSALLTYENAIHGFSTELTQERADSLMTQPGVIVLPSHR
YELHTRTPLEGLDEHTADLPFAGSYSDVVGVGLDTGVWPSKSYSDRGPRPISS
WKGCGEAGNTFASLGNKRLIGARFFARGYESTMGPIDESKSRPDDDHGHTHSS
TAAGSVVEGASLLGYAGSTGARGAPRAVAYKYVWLGCFSSDILAAIDKAIADNVN
VLSMGLGGMSDYIYRGDAVTCATFAAMEGILVCSAGNAGPSSGSLNAPWTTTGVGA
GLTRDDPALAILNGDVAITGVSLFKGEALPKDLPTIYAGNASNATNGNLCWMTGLI
PEKVKGIWQCDGINARVQGVKAGGVGMILANTANGBELVADAHLLPATTVG
EKAGDIRHVITDPTNPATISILGTVGVKPSVVAAFSSRGFNSITPNLLKPDLLA
PGVNIILAAWTGAAGTGLASDRRVEFNILSGTSMSCFHVSGLAALLKYHPEWSPA
IRALMTATYKTYKDGKPLDIAITGKSTPFDHGAGHVSPTTATNPGLIYDLTETDYL
GFCALNITYSPQIRSVRRNNTCDPSKSYADLVNPSFANVDGVGAYKTKTVTSV
WAGATYSKVTSITGVKIVSEPAVLNFKEANCKSYTVTFITVDSSKPSGSGNSFGSIE
MSDGKHYVGGSPVALSWI"
2275. .2305
3'UTR
BASE COUNT 547 a 567 c 580 g 611 t
ORIGIN
Alignment Scores:
Pred. No.: 2,93e-99 Length: 2305
Score: 1666.50 Matches: 362
Percent Similarity: 61.97% Conservative: 122
Best Local Similarity: 46.35% Mismatches: 246
Query Match: 41.48% Indels: 51
DB: 8 Gaps: 18

US-09-806-767-2 (1-775) x AY142612 (1-2305)
QY 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLys 25
Db 31 TTCTTCTC---CTCCCTTGTCTAGGATCTGCCACGTGCTCTCTCTCTCGACCAA 87
QY 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThr---PheAlaSer 44
Db 88 GGAACCTTACATCGTTTCATG-----GCTAAATCTCAGATGCGCTCG 129
QY 45 LysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGlu 64
Db 130 TCTTTCACCTCCACTTAATCGTATGATTCATCTCTCAGATCTATCTCCGAC----- 183
QY 65 GluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAla 84
Db 184 -----TCAGCCGAATTGCTCTACACTACAGAACCGGATTCATGATTCTCAACA 234
QY 85 GlnLeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArg 104
Db 235 CGGCTAACTCAAGAAGCCGACTCGCTCATGACTCAACTGGTGTTATCTCCGTTTAA 294
QY 105 ProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGly 124
Db 295 CCGGAGCACCGTTTACGAGCTACACACCACTCGTACTCTCTCTCTCTCGTCTCGACGAA 354
QY 125 -----PheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGly 139
Db 355 CACACCGCAGATCTGTTTCCCTGAAGCCGGCTCTTACAG-----GAC 396
QY 140 ThrIleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThr 159
Db 397 GTGCTGCTCGAGTGTCTCGATACCGGAGTTTGCCCGAGAGTAAAGCTACTCTCGACGAA 456
QY 160 GlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSer 179

```

QY 537 ProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysPro 556
 Db 1579 GCTTCGATCTCGCGCGGTGAGTTCAATATCTCTCGGACGTCGATGTTGGCCCT 1638
 QY 557 HisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaAla 576
 Db 1639 CAGGTAGTGGTTAGCGCGGTTCTCAAGTGGTGCATCTCGATGAGCCGCGCGG 1698
 QY 577 IleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyAlaIleLys 596
 Db 1699 ATTAGATCGCGGCTTATGACCGCTTACAAACCTACAAAGCGGTAAACCGTTACT 1758
 QY 597 Asp-----GlyAsnLysProAlaGlyValPheAlaIleGlyValAlaGlyHisValAsnPro 614
 Db 1759 GACATCGGACGAGGGAAGCTTCGACCGCTTCGATCAGCGTGCAGGACACGCTGCACCA 1818
 QY 615 GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr 634
 Db 1819 ACACCTGCCACTAATCCAGACTCATCTAGATCTAACGCGGAGATTAAGGCTTC 1878
 QY 635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer 654
 Db 1879 CTCTGCGCATTAATACACATCGCGCAATTCGAAGTGTCTCGAGACGTAATTACACT 1938
 QY 655 CysAsnGlyIleLeuArgLysAsnProGlyPheSer-----LeuAsnTyrProSer 671
 Db 1939 TGGCAT-----CCGAGTAATCGTACTCGTGGTGGTGAATTAACCTACCCGTCG 1986
 QY 672 IleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThrAsn 691
 Db 1987 TTGCGCGTTAACTGGAT---GGAGTCGGTGGGTATTAAGTACACGCGGACGAGTACGAGC 2043
 QY 692 ValGlySerProAsnSerIleTyrSerValAsnValLysAla---ProGluGlyIleLys 710
 Db 2044 GTGGGAGGAGCTGGACT---TACTCGGTTAAGTAAGTACCTCGGAGACGACGAGTCAAG 2100
 QY 711 ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg 730
 Db 2101 ATTTCGTTGAACCGCGGTTTGAATTTCAAGGAGTAAACGAGAAAAATCGTATACG 2160
 QY 731 ValTrpPheValLeuLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGly 750
 Db 2161 GTGACGTTTACTGTAGACTCGTCAAGCGCTCTGGA-----TCTAACGCTTTGGG 2211
 QY 751 GlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
 Db 2212 AGTATTGAATGGTCGATGGAGAAACACGTG-----GTGGGAATCCCGTGGCGATT 2262
 QY 771 Thr 771
 Db 2263 AGC 2265
 RESULT 14
 AF360285
 LOCUS Arabidopsis thaliana putative subtilisin serine protease ARA12
 DEFINITION (At5g67360) mRNA, complete cds.
 ACCESSION AF360285
 VERSION AF360285.1 GI:13430745
 KEYWORDS FLI CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 2618)
 YAMADA, K., LIU, S.X., SAKANO, H., PHAM, P.K., BANH, J., CHUNG, M.K.,
 GOLDSMITH, A.D., LEE, J.M., QUACH, H.L., TORIUMI, M., YU, G., BOWSER, L.,
 CARNINCI, P., CHEN, H., CHEUK, R., HAYASHIZAKI, Y., ISHIDA, J.,
 JONES, T., KAMIYA, A., KARLIN-NEUMANN, G., KAWAI, J., KIM, C., LAM, B.,
 LIN, J., MIRANDA, M., NARUSAKA, M., NGUYEN, M., PALM, C.J., SAKURAI, T.,
 SATOU, M., SEKI, M., SHINN, P., SOUTHWICK, A., SHINOZAKI, K.,
 DAVIS, R.W., ECKER, J.R. and Theologis, A.

Davis, R.W., Ecker, J.R. and Theologis, A.
 Arabidopsis Full Length cDNA Clones

TITLE JOURNAL REFERENCE AUTHORS

2 (bases 1 to 2618)
 Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K.,
 Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
 Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
 Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
 Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
 Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
 Davis, R.W., Ecker, J.R. and Theologis, A.

TITLE JOURNAL

Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAPL cDNAs (RAPL cDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
 sequencing and annotation of the RAPL cDNAs: Yamada, K., Liu, S.X.,
 Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D.,
 Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Chen, H.,
 Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J.,
 Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A.,
 Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC)
 contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
 genome submitted to Genbank.

FEATURES source

1..2618
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="RAPL05-16-L13 (R09277)"
 /note="This clone is in a modified pBluescript vector
 (F1C-1) as a BamHI/XhoI insert."
 ecotype="Columbia"

gene 5'UTR CDS

1..2618
 /gene="At5g67360"
 1..114
 /gene="At5g67360"
 115..2388
 /gene="At5g67360"
 /note="contains subtilase domains PF00082 and a PA domain
 PF02225"
 /codon_start=1
 /evidence=experimental
 /product="putative subtilisin serine protease ARA12"
 /protein_id="AAK25995.1"
 /db_xref="GI:13430746"
 /translation="MSSSFLSSTAFFLLCLGFCVSSSSDQGTIVHMAKSMQPS
 FDLNSWYDSSLRISDSABLLTYENAIGHFSTRTOEADSLMTQPGVISVLPBHR
 YELHTRTPLEGLDEHDLDFEAGSYSDVGVVLDGTWPKSKSYSDGFGPIPS
 WKGCCEAGTFTASLNCRLKIGARFFARGFESTMGPIDESKSRSPDDDDGCHTHRS
 TAAGSVVVGASLIGYAGTAGMAFPAARVAVYKVCWGGCFSSDILAAIDKALADNVN
 VLMSLGGNGSYRDGVATGAFPAAMERGLVSCSAGNAGPSSLSNVAPWITTVGA
 GTLDKDPALAILNGKNFTGSLFKGELPKLLPFIYAGNASNATNGMLCMTGLI
 PEKVGKIVMCDGIRNARVCKGVKAAAGGVGMILANTRAANGELVADAHLLPATTVG
 EKAGDIIRHVITDPTASISILGTVGVPSPVVAAPSSRGPNSTNLPKPDLIA
 FGNVILAAWTGAAGPTGLASDRRVEFNIIISGSMSCPHVSGLAALLKSVHPWSPAA
 IFSALMTAYTKYDKGKPLDIDATGKSPFPDGHAGHVSPTATNPGLIYDLITEDYL
 GFLCANMTSPQIRSVSRNRYTCDPSKSYADLNTPSPFVAVNVDGAYKTYTIVSY
 GGAGTSYKVTSETTGVKISVEPAVLNFKEANEKSYTFTFTDSSKPSGSGNSFGSIE
 WSGGKHVGVSPVAISWT"
 2389..2618
 /gene="At5g67360"

3'UTR


```

Db 1993 CTCTGGCATTGAATTACATCGCGCAAAATTCGAAGTGTCTCGAGACGTAATTACACT 2052
Qy 655 CysAsnGlyIleuArgLysAsnProGlyPheSer-----LeuAsnTyProSer 671
Db 2053 TCGGAT-----CCGAGTAAATCGTACTCGTGCNGATTTGAACACTACCCGTCG 2100
Qy 672 IleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThrAsn 691
Db 2101 TTCGCCGTTAAGCTTCAT---GGATCGGTGCGGTATAGTACACGCGCAGCGTACGAGC 2157
Qy 692 ValGlySerProAsnSerIleTySerValAsnValLysAla---ProGluGlyLys 710
Db 2158 GTGGAGGAGCTGGGACT--TACTCGGTAAAGTAACTTCGGACACACAGGAGTCAAG 2214
Qy 711 ValIleValAsnProLysArgLysValPheLysHisValAspGlnThrLysSerTyArg 730
Db 2215 ATTTCGGTTGAACCGCGCGTTTGAATTTCAAGGAAGCTAACAGAAATAATCGTATACG 2274
Qy 731 ValTyrPheValLeuLysLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGly 750
Db 2275 GTACGCTTTACTGTAGACTCGTCGAAGCGCTCTGGA-----TCTAACAGCTTTGGG 2325
Qy 751 GlnLeuThrTyrValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
Db 2326 AGTATTGAATGTCGGATCGGAACACAGTG-----GTGGGAAGTCCCGTGGCGATT 2376
Qy 771 Thr 771
Db 2377 AGC 2379

RESULT 15
LOCUS AF065639 2686 bp DNA linear PLN 14-AUG-2000
DEFINITION Arabidopsis thaliana cucumis-in-like serine protease (ARAI12) gene,
complete cds.
ACCESSION AF065639
VERSION AF065639.1 GI:3176873
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 2686)
AUTHORS Ribeiro,A., Akkermans,A.D., van Kammen,A., Bisseling,T. and
Pawlowski,K.
TITLE A nucleotide-specific gene encoding a subtilisin-like protease is
expressed in early stages of actinorhizal nodule development
JOURNAL Plant Cell 7 (6), 785-794 (1995)
MEDLINE 95375542
PUBMED 7647567
REFERENCE 2 (bases 1 to 2686)
AUTHORS Yamagata,H., Uesugi,M., Saka,K., Iwaseki,T. and Aizono,Y.
TITLE Molecular cloning and characterization of a cDNA and a gene for
subtilisin-like serine proteases from rice (Oryza sativa L.) and
Arabidopsis thaliana
JOURNAL Biosci. Biotechnol. Biochem. 64 (9), 1947-1957 (2000)
MEDLINE 20506592
PUBMED 11053401
REFERENCE 3 (bases 1 to 2686)
AUTHORS Yamagata,H.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-1998) Laboratory of Biochemistry, Faculty of
Agriculture, Kobe University, Nada, Kobe 657-8501, Japan
FEATURES
source
location/Qualifiers
1..2686
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
<290..>2563
/gene="ARAI12"
<290..>2563

```

```

/gene="ARAI12"
/product="cucumis-in-like serine protease"
290..2563
/gene="ARAI12"
/notes="similar to subtilisin-like protein encoded by
GenBank Accession Number X85974; intronless gene"
/codon_start=1
/product="cucumis-in-like serine protease"
/protein_id="AAC18851.1"
/db_xref="GI:3176874"
/translation="MSSSFLSTAFLLCLGFCVHVSSSSSSQGYIIVMAKSMQPS
FELHNSYDSLSRSISDAELTYENAIHGFSTRLTQEEASLMTQPGVIVSLPEHR
YELHTRTFLFLGDEHATDLPEAGSYDVVGVLDGTGWPEKSYSGDEGFPVPS
WKGCGAGTNTASLCNRKLGARFARGYSTMGPIDESKESRPRDDDGHTHTSS
TAAGSVVEGASLLGVASGTARGVAPRVAVVYKVLGCGFSSDILAAADKALADNV
VLSMGLGGWSYDRDGVAGAFAMERGLIIVCSAGNAGPSSSLSNVAPITTVGA
GLTRDFPALAALNGNFTGVSFLFGEALPKLIPFIYAGNASNATNCLMCTGLI
PEKYGKIVMCDRGINARVQKGVVKAAGVGVMILANTAAAGEELVADHLIPATTVG
EKAGDIIRHVTTDNPFTASISILGTVGVKFPVVAFFSRGPNSITPNIILKPLDIA
PGVNIILAAATGAAGPTGLASDSRRVEFNIISGSMSCPHVSGLAALLKSVHPWSPAA
IRSLMTTAYKYKQKGPLLDIATGKPSPTPDHAGHVSPTTATNPGLYDLITDYL
GFULALNTYSPQISVSRNNYTCDPKSYSVADLNPSPFVNVVDGVGAYKRYTRTVSV
GGATYSVKVTSSTTGGIKISVPEVLNFKAEKKSYTVTFTVDSKSPGSGNSFGSIE
WSDGKHVQSPVAISWT"
BASE COUNT 687 a 646 c 639 g 714 t
ORIGIN
Alignment Scores:
Pred. No.: 3,62e-99 Length: 2686
Score: 1666.50 Matches: 362
Percent Similarity: 61.97% Conservative: 122
Best Local Similarity: 46.35% Mismatches: 246
Query Match: 41.48% Indels: 51
DB: 8 Gaps: 18
US-09-806-767-2 (1-775) x AF065639 (1-2686)
Qy 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLys 25
Db 320 TTTCTTCTC---CTCCTTTGCTAGGATTCCTGCACGTGCTCTCTCCCTCCGACCAA 376
Qy 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThr---PheAlaSer 44
Db 377 GGAATTTACATCGTTTCAATG-----GCTAAATCTCAGATCCGCTCG 418
Qy 45 LysPheAspTyrPheHisLeuSerPheLeuGlnLysValLeuGlyValGluGlnGlu 64
Db 419 TCTTTGACCTCCACTTCACTGTCAGATTCTATCTCTCAGATCTATCTCCGAC----- 472
Qy 65 GluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAla 84
Db 473 -----TCAGCGCAATTGCTCTACACTTACGAGACGCGATTTCATGATTCTCAACA 523
Qy 85 GlnLeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArg 104
Db 524 CGGTAACTCAAGAAGCGCGACTCGCTCATGACTCAACCTGCTGTTATCTCCGTTTAA 583
Qy 105 ProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGly 124
Db 584 CCGGAGCACCGTTACGAGCTACACACACTCTGTAATCTCTCTCTCTCTCGTCTCGAGAA 643
Qy 125 -----PheGlyAsnSerGlyValTyrSerLysSerArgPheGlyGlnGly 139
Db 644 CACACCGCAGATCTGTTCTCCCTGAAGCGCGCTCTTACAGC-----GAC 685
Qy 140 ThrIleIleGlyValLeuAspThrGlyValTyrProGluSerProSerPheAspAspThr 159
Db 686 GTCTGCTCGGAGTCTTCGATACCGGAGTTGGCCGAGAGTAAAGTAAAGTAAAGTAAAG 745
Qy 160 GlyMetProSerIleProArgLysTyrLysGlyIleCysGlnGluGlyGluSerPheSer 179
Db 746 GGATTCTGCTCTATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 805

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 19:57:04 ; Search time 445 Seconds
(without alignments)
14090.340 Million cell updates/sec

Title: US-09-806-767-1

Perfect score: 2328

Sequence: 1 atggaacccaacacctttctt.....taacctgaagactaactga 2328

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2328	100.0	2328	21	AAZ93808
2	2326.4	99.9	2328	21	AAZ93809
3	1024.6	44.0	2492	21	AAZ93813
4	1020.2	43.8	3140	21	AAZ93814
5	602	25.9	3855	21	AAZ93812
6	481.8	20.7	2265	21	AAZ93317
7	481.8	20.7	2265	24	ABZ13260
8	447.4	19.2	2434	21	AAZ47609

9	424.4	18.2	2637	21	AAZ46158
10	417	17.9	2638	21	AAZ42012
11	398.2	17.1	2295	24	ABZ13381
12	398.2	17.1	2489	21	AAZ42230
13	285	12.2	1653	21	AAZ32938
14	228.2	9.8	2402	21	AAZ42099
15	174.6	7.5	740	24	ABQ65482
16	168.2	7.2	2552	15	AAQ73756
17	123.2	5.3	436	25	ABX61910
18	121.2	5.2	2310	25	ABZ42107
19	113.6	4.9	2448	20	AAZ29905
20	110.4	4.7	975	21	AAZ49798
21	100.2	4.3	2958	24	ABZ12533
22	90.8	3.9	513445	22	AAI61373
23	83.4	3.6	363	25	ABX20410
24	82	3.5	7235	20	AAZ29910
25	78.8	3.4	343	25	ABX21785
26	71.2	3.1	279	25	ABX23353
27	67.8	2.9	401	25	ABX19355
28	65	2.8	513445	22	AAI61373
29	63	2.7	528	25	ABX57523
30	62.4	2.7	297	24	ABL76753
31	62.4	2.7	379	25	ABX20319
32	58	2.5	293	25	ABX88421
33	55.8	2.4	2451	24	ABZ12259
34	51	2.2	255	25	ABX34387
35	49	2.1	273	24	ABL73078
36	44.8	1.9	1236	20	AAZ05920
37	44.8	1.9	1566	18	AAZ85668
38	44.8	1.9	1962	18	AAZ85695
39	44.8	1.9	1962	20	AAZ05929
40	44.4	1.9	295	25	ABX88947
41	43.6	1.9	454	24	ABL33821
42	42.6	1.8	4590	22	AAZ24065
43	41.6	1.8	531	25	ABX57480
44	41.6	1.8	272	24	ABL72301
45	40.2	1.7	569	25	ABX56955

ALIGNMENTS

RESULT 1

AAZ93808

ID AAZ93808 standard; DNA; 2328 BP.

XX AAZ93808;

XX AAZ93808;

DT 16-AUG-2000 (first entry)

DE SDD1 sequence encoding a subtilisin like serine protease.

XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;

KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;

KW crop protection; feed; foodstuffs; ss.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Berger D, Altmann T;
 XX WPI: 2000-317995/27.
 DR P-PSDB; AAY83300.
 XX
 PT Novel recombinant DNA molecules encoding subtilisin-like serine
 PT protease, useful for producing transgenic plants with altered stomata,
 PT lower water consumption and enhanced diseased resistance
 XX
 XX Claim 1; Page 60-63; 101pp; English.
 XX
 CC Sequences encoding SDP1, a subtilisin-like serine protease, can be
 CC used to produce transgenic plants with altered stomata
 CC characteristics. These plants exhibit improved freshness,
 CC increased dry weight, reduced leaf temperatures, reduced water loss
 CC and lower water consumption and for enhancing the sugar and/or
 CC protein content of plant leaves, modulating CO₂ uptake into and H₂O
 CC release from leaves, for sustained photosynthesis under high
 CC intensity conditions or for the improvement of disease resistance
 CC of plants. The transgenic plants and cells of such plants are useful
 CC in the preparation of feed, food or additives.
 XX
 XX Sequence 2328 BP; 635 A; 488 C; 577 G; 628 T; 0 other;
 SQ
 Query Match 100.0%; Score 2328; DB 21; Length 2328;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGAACCCAAACCTTCTCTCTGCAATATCTTCTCTATTTCTTCTTCTTCTGCA 60
 DB 1 ATGGAACCCAAACCTTCTCTCTGCAATATCTTCTCTATTTCTTCTTCTTCTGCA 60
 QY 61 GAGATCCTGAGAGAGAGACTTACATGTTTCTGATTCATCTTCTGATTCATCTGATAA 120
 DB 61 GAGATCCTGAGAGAGAGACTTACATGTTTCTGATTCATCTTCTGATTCATCTGATAA 120
 QY 121 ACCTTTGCCCTCAAGTTTGAATGGCATCTTCTTCTTCTGATTCATCTTCTGATTCATCT 180
 DB 121 ACCTTTGCCCTCAAGTTTGAATGGCATCTTCTTCTTCTGATTCATCTTCTGATTCATCT 180
 QY 181 GAAGAAGAGAGAGAGAGAGCTTCTTCTGATTCATCTTCTGATTCATCTTCTGATTCATCT 240
 DB 181 GAAGAAGAGAGAGAGAGAGCTTCTTCTGATTCATCTTCTGATTCATCTTCTGATTCATCT 240
 QY 241 GGATTTGCTGCTCAGTTGACTGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 DB 241 GGATTTGCTGCTCAGTTGACTGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 QY 301 GTTGCACTGAG 360
 DB 301 GTTGCACTGAG 360
 QY 361 GGACTGACAG 420
 DB 361 GGACTGACAG 420
 QY 421 ATTATCGGCGTGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 DB 421 ATTATCGGCGTGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 QY 481 ATGCTTTCGATTCACGAG 540
 DB 481 ATGCTTTCGATTCACGAG 540
 QY 541 TCGAGCTGTAAACCGAG 600
 DB 541 TCGAGCTGTAAACCGAG 600
 QY 601 AATTCCACGAG 660
 DB 601 AATTCCACGAG 660

QY 661 GGACACGGGACTCACACCGGCTCAACAGTTGTTGGATCTCTGTTTTCGATGGCGAATGTT 720
 DB 661 GGACACGGGACTCACACCGGCTCAACAGTTGTTGGATCTCTGTTTTCGATGGCGAATGTT 720
 QY 721 CTTGGCAATGAGCTGTTGGCTCTGGATGGCTCTGGAGTGGCTCTGGAGCTCACATTCAGTCTAT 780
 DB 721 CTTGGCAATGAGCTGTTGGCTCTGGATGGCTCTGGAGTGGCTCTGGAGCTCACATTCAGTCTAT 780
 QY 781 AAAGTCTGTTGGTTCAATGGTCTTACAGCTCTGACATTCAGCTCTAGAGCTATAGATGAGCG 840
 DB 781 AAAGTCTGTTGGTTCAATGGTCTTACAGCTCTGACATTCAGCTCTAGAGCTATAGATGAGCG 840
 QY 841 ATTCAAGATAAAGTCGATGTTTTCGCTTCCCTTGGCGGTTTCCCTATTCCTTTGAT 900
 DB 841 ATTCAAGATAAAGTCGATGTTTTCGCTTCCCTTGGCGGTTTCCCTATTCCTTTGAT 900
 QY 901 GATGACACAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 960
 DB 901 GATGACACAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 960
 QY 961 GCAGCTGTTAAACAGGTCCTCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 1020
 DB 961 GCAGCTGTTAAACAGGTCCTCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCG 1020
 QY 1021 ACCATTGGCGCAGCAGCTTGTATCGAAGATTTCCGCTGTGTGTCAGATTAGCCACGGA 1080
 DB 1021 ACCATTGGCGCAGCAGCAGCTTGTATCGAAGATTTCCGCTGTGTGTCAGATTAGCCACGGA 1080
 QY 1081 AAGCTTCTCTATGAGAGTCAATGTTATCCGGGAAAAGGTATTAAGAAATCCCGGAGAGAG 1140
 DB 1081 AAGCTTCTCTATGAGAGTCAATGTTATCCGGGAAAAGGTATTAAGAAATCCCGGAGAGAG 1140
 QY 1141 GTTGAGTGTATTTACGTCACAGAGAGAGATTAAGAAATTCGCTGTGTGTCAGAGGTC 1200
 DB 1141 GTTGAGTGTATTTACGTCACAGAGAGAGATTAAGAAATTCGCTGTGTGTCAGAGGTC 1200
 QY 1201 CTTCCAAAGAGAGAGATTCGAGGCAAAATCGTGATTTGTGATCGCGAGTCAATGGAAGA 1260
 DB 1201 CTTCCAAAGAGAGAGATTCGAGGCAAAATCGTGATTTGTGATCGCGAGTCAATGGAAGA 1260
 QY 1261 TCGGAGAAAGAGAGAGAGCGGTAAAGAGAGCTGAGAGAGTTCGCAATGATCTTATAGCAATACA 1320
 DB 1261 TCGGAGAAAGAGAGAGAGCGGTAAAGAGAGCTGAGAGAGTTCGCAATGATCTTATAGCAATACA 1320
 QY 1321 GAGATCAACCAAGAGAGAGATTTATGAGCTTCATCTCTTACAGCTACATTTGATTTG 1380
 DB 1321 GAGATCAACCAAGAGAGAGATTTATGAGCTTCATCTCTTACAGCTACATTTGATTTG 1380
 QY 1381 TACACTGAGTCACTCTTCTGAGGCTTATGTTAATGCCACGCTGAAACCAAGAGCGCG 1440
 DB 1381 TACACTGAGTCACTCTTCTGAGGCTTATGTTAATGCCACGCTGAAACCAAGAGCGCG 1440
 QY 1441 ATAAATTTTGGTGTACCGTGTATGGAGGTCACAGAGACCGGAGGTGCTCAGTTTTC 1500
 DB 1441 ATAAATTTTGGTGTACCGTGTATGGAGGTCACAGAGACCGGAGGTGCTCAGTTTTC 1500
 QY 1501 GCTCAGAGACCGAGTTTATGCAATTCCTTATGATTAATAAACCAGATATGATTCGCGGA 1560
 DB 1501 GCTCAGAGACCGAGTTTATGCAATTCCTTATGATTAATAAACCAGATATGATTCGCGGA 1560
 QY 1561 GTCAATATCATTTGGGCTTGGCTTCAAAATCTAGGACCAACCGGACTTCCTTATGATTC 1620
 DB 1561 GTCAATATCATTTGGGCTTGGCTTCAAAATCTAGGACCAACCGGACTTCCTTATGATTC 1620
 QY 1621 AGAAGAGTTAACTTCACTGTAAATGTCAGGAACCTTCAATGCTTGTCCACATGTTAGCGGA 1680
 DB 1621 AGAAGAGTTAACTTCACTGTAAATGTCAGGAACCTTCAATGCTTGTCCACATGTTAGCGGA 1680
 QY 1681 ATCACTGCTCTTATTCGCTGTGATACCCGAGTGTCTCCAGTGTCAATCAATTCGCA 1740
 DB 1681 ATCACTGCTCTTATTCGCTGTGATACCCGAGTGTCTCCAGTGTCAATCAATTCGCA 1740
 QY 1741 TTGATGACACAGCGGATTTGTGATGCTCAAGGGGAAACCGGATAAAGGATGTTACAA 1800

601 AATTCCAGAGAGAAATCAACAAACATGCTCTGTGAATACATTTCCGCAAGAGATTCAACG 660
661 GGACAGGGAGCTCACACCGGCTCAACAGTTGGTGGATCCTCTGTTTCGATGCGGAATGTT 720
661 GGACAGGGAGCTCACACCGGCTCAACAGTTGGTGGATCCTCTGTTTCGATGCGGAATGTT 720
721 CTTGGCAATGAGAGCTGGTGGTCTGTGGATGGCTCTGGAGCTCAACATTCAGTCTAT 780
721 CTTGGCAATGAGAGCTGGTGGTCTGTGGATGGCTCTGGAGCTCAACATTCAGTCTAT 780
781 AAGTCTGTGGTCAATGGTGTGTACAGCTCTGACATCTAGCAGCTATAGATGAGCG 840
781 AAGTCTGTGGTCAATGGTGTGTACAGCTCTGACATCTAGCAGCTATAGATGAGCG 840
841 ATTCAAGATAAAGTCAGATGTTCTTTCGCTTCCCTTGGCGGTTTCCCTATTCTCTTGTAT 900
841 ATTCAAGATAAAGTCAGATGTTCTTTCGCTTCCCTTGGCGGTTTCCCTATTCTCTTGTAT 900
901 GATGACACATCCCATTTGGAACATTCGAGGCCATGGAACGCGGTATATCTGTAATCTGT 960
901 GATGACACATCCCATTTGGAACATTCGAGGCCATGGAACGCGGTATATCTGTAATCTGT 960
961 GCAGCTGTGTAACAAACGGTCCATCGAAAGCTCTGTTGCAAAACACAGCTCTTGGGTCTCA 1020
961 GCAGCTGTGTAACAAACGGTCCATCGAAAGCTCTGTTGCAAAACACAGCTCTTGGGTCTCA 1020
1021 ACCATTGGCGCAGGACGCTTGATCGAGATTTCCCGCTGTGGTCAAGATTAGCCAAACGGA 1080
1021 ACCATTGGCGCAGGACGCTTGATCGAGATTTCCCGCTGTGGTCAAGATTAGCCAAACGGA 1080
1081 AAGCTTCTCTATGAGAGTCAATGTTATCCGGGAAAAGGTATAAAGAAATCCGGGAGAGAG 1140
1081 AAGCTTCTCTATGAGAGTCAATGTTATCCGGGAAAAGGTATAAAGAAATCCGGGAGAGAG 1140
1141 GTTGAGGTGATTTACGTCAGAGAGAGATTAAGAAAGTGAAGTCTGTTTGAAGAGGTCA 1200
1141 GTTGAGGTGATTTACGTCAGAGAGAGATTAAGAAAGTGAAGTCTGTTTGAAGAGGTCA 1200
1201 CTTTCAAGAGAAGAAATCCGAGCAAAATGGTGAATTTGATCGCGAGTCAATGGAAGA 1260
1201 CTTTCAAGAGAAGAAATCCGAGCAAAATGGTGAATTTGATCGCGAGTCAATGGAAGA 1260
1261 TCGGAGAAAGAGAGCGGTTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACA 1320
1261 TCGGAGAAAGAGAGCGGTTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACA 1320
1321 GAGATCAACCAAGAAGAAGATCTATTGAGCTTCACTCTTACCAGCTACATTGATTGTT 1380
1321 GAGATCAACCAAGAAGAAGATCTATTGAGCTTCACTCTTACCAGCTACATTGATTGTT 1380
1381 TACACTGAGTCAGTCCCTCTGAAGGCTTATGTTAATGCCACCGTGAACCCAAAGGCGGG 1440
1381 TACACTGAGTCAGTCCCTCTGAAGGCTTATGTTAATGCCACCGTGAACCCAAAGGCGGG 1440
1441 ATAAATTTTGGTGTACGGTGTATGGAGGTACAGAGCACCGAGGTGCTCAGTTTCA 1500
1441 ATAAATTTTGGTGTACGGTGTATGGAGGTACAGAGCACCGAGGTGCTCAGTTTCA 1500
1501 GCTCGAGGACCGAGTTTAGCCATCTCTGATATAAAACCGGATATGATGCTCCGGGA 1560
1501 GCTCGAGGACCGAGTTTAGCCATCTCTGATATAAAACCGGATATGATGCTCCGGGA 1560
1561 GTCAATATCATTTGCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTC 1620
1561 GTCAATATCATTTGCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTC 1620
1621 AGAAGATTAACTTCACTGTAATGTCAAGAACCTCAATGCTTGTCCACATGTTAGCGGA 1680
1621 AGAAGATTAACTTCACTGTAATGTCAAGAACCTCAATGCTTGTCCACATGTTAGCGGA 1680
1681 ATCACTGCTCTTATCCGCTGTCATACCCGAACCTGCTTCCAGCTCAATCAAAATCCGGA 1740
1681 ATCACTGCTCTTATCCGCTGTCATACCCGAACCTGCTTCCAGCTCAATCAAAATCCGGA 1740

1741 TTGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAGCGGATAAAGATGTTAAACAA 1800
1741 TTGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAGCGGATAAAGATGTTAAACAA 1800
1801 CCAGCCGGTGTGTTCCGATTTGAGCAGGCGCATGTGAATCCGCAAAAGCGGATAAACCAG 1860
1801 CCAGCCGGTGTGTTCCGATTTGAGCAGGCGCATGTGAATCCGCAAAAGCGGATAAACCAG 1860
1861 GGATTTGTTTACAAACATTTCAACAGTGGATTACATAAATTTACCTTGCACCTTTGGATT 1920
1861 GGATTTGTTTACAAACATTTCAACAGTGGATTACATAAATTTACCTTGCACCTTTGGATT 1920
1921 ACAGATCAGATATTTTAGCAATCACTCATPAAGAACGTGAGCTGCAATGGAATATTGCGG 1980
1921 ACAGATCAGATATTTTAGCAATCACTCATPAAGAACGTGAGCTGCAATGGAATATTGCGG 1980
1981 AAAAACCCGGGTTTAGTCTCAANTTACCCTCGATAGCCGTGATTTTCAACCTGGCAAG 2040
1981 AAAAACCCGGGTTTAGTCTCAANTTACCCTCGATAGCCGTGATTTTCAACCTGGCAAG 2040
2041 ACTACGGAGATGATCAACAGCGGTGTCACATAACGTTGGGAGTCTTAACCTCGATATACTCA 2100
2041 ACTACGGAGATGATCAACAGCGGTGTCACATAACGTTGGGAGTCTTAACCTCGATATACTCA 2100
2101 GTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCAATCCTAAGAGACTTGTGTT 2160
2101 GTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCAATCCTAAGAGACTTGTGTT 2160
2161 AAACACCTGTGATCAGAGCTGAGCTATAGAGTATGGTTTGTATTTGAAGAGAAACAG 2220
2161 AAACACCTGTGATCAGAGCTGAGCTATAGAGTATGGTTTGTATTTGAAGAGAAACAG 2220
2221 GGAGGGAAGTGGCTAGCTTTGACAGAGGCGAGTGGCTGGCTCACTCTCAATCTG 2280
2221 GGAGGGAAGTGGCTAGCTTTGACAGAGGCGAGTGGCTGGCTCACTCTCAATCTG 2280
2281 ATGCAGCGAGTTAGAAGTCCAATCTCTGTAACTTGAAGACTAACTGA 2328
2281 ATGCAGCGAGTTAGAAGTCCAATCTCTGTAACTTGAAGACTAACTGA 2328

RESULT 3

AAZ93813
ID AAZ93813 standard; DNA; 2492 BP.
XX AC AAZ93813;
XX 16-AUG-2000 (first entry)
XX Subtilase (Pigesp) of Solanum tuberosum.
XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
XX stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
XX crop protection; feed; foodstuffs; ss.
XX Solanum tuberosum.

OS Solanum tuberosum.
XX Key Location/Qualifiers
FH CDS 3..2492
FT /*tag= a
FT /product= Subtilase Pigesp
XX WO200022144-A2.

PN PD 20-APR-2000.
XX PF 12-OCT-1999; 99WO-EP07633.
XX PR 12-OCT-1998; 98EP-0119244.
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

XX Berger D, Altmann T;
XX WPI; 2000-317995/27.
XX P-PSDB; AAY83303.
XX Novel recombinant DNA molecules encoding subtilisin-like serine
XX protease, useful for producing transgenic plants with altered stomata,
XX lower water consumption and enhanced diseased resistance
XX Claim 1; Page 79-82; 10pp; English.
XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
XX used to produce transgenic plants with altered stomata
XX characteristics. These plants exhibit improved freshness,
XX increased dry weight, reduced leaf temperatures, reduced water loss
XX and lower water consumption and for enhancing the sugar and/or
XX protein content of plant leaves, modulating CO₂ uptake into and H₂O
XX release from leaves, for sustained photosynthesis under high
XX intensity conditions or for the improvement of disease resistance
XX of plants. The transgenic plants and cells of such plants are useful
XX in the preparation of feed, food or additives. This subline of
XX Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
XX (See AA293808).
XX Sequence 2492 Bp; 766 A; 480 C; 511 G; 734 T; 1 other;
Query Match 44.08; Score 1024.6; DB 21; Length 2492;
Best Local Similarity 68.18; Pred. No. 5.1e-309;
Matches 1455; Conservative 0; Mismatches 675; Indels 6; Gaps 2;
QY 189 AGAGGAGAGCCCTCTCTCGACTTCTCTACTCTCTATGGCTCGCGATTGAAGGATTGC 248
DB 356 AGGTGAAAATCGAGTTCTCGCTTTTGTTGTTACTCTTACCATTCTGCATTGAGAGTTTGC 415
QY 249 TGCTCAGTTGACTGAATCAGAAGCCGAGATACATGAGATATTCACCTGAAAGTTGTTGCAGT 308
DB 416 AGCACTTCTATCGAAATGAGCTAAAGGCACCTGAAGAAATCGAATAATGTGTATCAAT 475
QY 309 GAGACCTGACATGTTCTTCAGTTTCAACCACTTACTCTTACAAGTTCTTGGAGCTCGA 368
DB 476 ATATCCGGAGAGGAAGCTTGAGGTTCAAAACAATTATCTTACAAGTTCTTAGGACT--- 532
QY 369 CGGTTTGGAACTCCGGTGTATGGTCTAATCTCGTTTGGTTCAGGACATATATCGG 428
DB 533 TAGTCTTACAAAGGAAGTACTTGGTTAAAGTCTGAGTTGGTTCAGGCGCGCATCTGG 592
QY 429 CGTCTTGTACTGAGTTTGGGCTGAAAGTCTAGCTTGTGACATACCGGAATGCTTTC 488
DB 593 AGTCTTGTACTGGAATTTGGCAGAAAGTCCAAAGTTTGTGATCATGGAAATGCTTCC 552
QY 489 GATTCCACGGAATGGAAGGGATTGCGCAAGAGGAGAAAGTTTCAGTTCTTTCGAGGTG 548
DB 653 TATTCCAAAGAAATGGAAGGTTNTGCGCAAGAGGAAAGAACTTCAATTTCTCAAGTTG 712
QY 549 TAACCGGAAGTAAATCGGTGCTAGATTTCTTCATCAGAGGACCGGTGCTGCTAATTCACC 608
DB 713 CAATTCGCAAGCTTATTGTTGTCAGAGTTTTCAGATAGGACACATGATGCGCATCAAGAC 772
QY 609 AGAGGATACCAAAACATGCTCGTGAATACATTTCCGCAAGAGATTCAACGGACACGG 668
DB 773 ATCAAATCAATAGATTTATGAGGAGATTATGATCATCTCGAGATTCTCAAGGCGCATGG 832
QY 669 GACTTCACACCGCTCAACAGTTGGTGGATCTCTGTTTGGATGGCGGAATGTTCTTGGCAA 728
DB 833 TACACATACAGCATCTACTCAGGGGAGCTCCCGTTCCAAATGCGGAGTGTCTTGGAAA 892
QY 729 TGGAGCTGGTGTGGCTCGTGGATGGCTCTGAGAGTCAATTCAGTCTATAAAGTCTG 788
DB 893 TGGAGCAGGAGAGGCTCGAGGATGGCCCTGGTGTCTATCGCATATACAAAGTTTG 952
QY 789 TTGTTTCAATGGTTGTTACAGTCTGACATTTCTAGAGCTATAGATGAGCGATTCAAGA 848

DB 953 TTGCTCTAGTGGTTGTATAGTTCTGATATACCTTGCAGCAATGGATAGCTATTAGAGA 1012
QY 849 TAAAGTCGATGTTCTTTCGCTTTCCTTGGCGGTTTCCCTATTCTTCTTGTATGATGACAC 908
DB 1013 TGGAGTAGACATATGTCTCTTTCAATTTGGTGGTTTCCCTGTTCCATTTATGAGGATAC 1072
QY 909 AATCGCCATTGGAACATTCGAGCCATGGAACCGGTATATCTGTATCTGTGAGCTGG 968
DB 1073 TATTGCTATTGGCAGTTTTCGAGCTATGGAACGTGGAATTTCAATTTATGTTGCTGAGG 1132
QY 969 TAAACAACGGTCCAAATCGAAAGCTCTGTTGCAACACAGCTCTTGGGTCTCAACCATGG 1028
DB 1133 AAATAAGTGTCCAAATCTTAAGTTTCAGTAGCAAAATGAGGCTCTTGGATTGCCATATTGG 1192
QY 1029 CGCAGGACAGCTTGCATCGAAGATTTCCGCTGTGGTCAGATTAGCAACGAAAGCTTCT 1088
DB 1193 TGCTAGCACATTTGACAGGAAATTTCCAGCAATTAATTCAGCTAGGTAAATGGCAAGTATGT 1252
QY 1089 CTATGGAGAGTCATTTGTATCCGGAAAAAGGTATAAAGAAATGCCGGGAGAGAGGTGAGGT 1148
DB 1253 GTATGGAGAAATCTTGTACCGGGCAAAACAAGTTCAATAATCTCAGAAAAGTTCTTGAGAT 1312
QY 1149 GATTTAGCTACAGAGAGAGATAAAGGAAGTGTCTTCTGTTGAGAGGCTCACTTCCAAG 1208
DB 1313 TGTTTATCTCAATGACGGTGTGATAATGGAAGTGAATTTTGCTTAAGAGGGTCTCTGCCAAG 1372
QY 1209 AGAAGAAATCCGAGGCAAAATGTTGATTGTCGCGGAGTCAATGGAAGTCGGAGAA 1268
DB 1373 AGCTAAAGTCCATGGAATAATTCGTTGTATGTGATCGTGGAGTTAATGGAAGAGCAGAGAA 1432
QY 1269 AGGAGAAGCGGTTAAAGAAAGCTGGAGGAGTTGCAATGATCTTAGCCAAATCAGAGATCAA 1328
DB 1433 AGGTCAAGTTGTTAAAGAAATCAGTGGTGTGCTCATGATCTTAGCAATACAGCAGTAAA 1492
QY 1329 CCAAGAAAGAGATTCTATGACGTTCACTCTTACCAGCTACATGTATGTTGTACACTGA 1388
DB 1493 TATGAGGAAAGATTCTGTGGACGTACATGTCTACCTGCAACATTTGTTTGGACGA 1552
QY 1389 GTCACTCTTCTGAAGGTTATGTTAATGCCACGGTGAACCAAGCGCGGATTAATTTT 1448
DB 1553 ATCAATTCAGTTGCAAAAGCTATATGAATCAACCGCAAAACCAACAGCTCGAATCATATT 1612
QY 1449 TGGTGTACGGTGTGATTTGGAGGTCACGAGCACCGAGGTGCTCAGTTTTCAGCTCCAGG 1508
DB 1613 TGGAGAAACAGTTATAGAAAATCTAGTGACCTGCTGTAGCAAAATTTTCTTCAAGGG 1672
QY 1509 ACCGAGTTTACCCAAATCCTTCGATACATAAAACCGGATATGATTGCTCCGGGAGTCAATAT 1568
DB 1673 TCCAGTTTCTACTGATCTTCAATTTCTCAAACTGTATGTTGCTCCAGGTGTCAACAT 1732
QY 1569 CATTCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTCTTATGATTCAAGAAAGT 1628
DB 1733 AATTGCTGTTGGCGGCAAAATCTAGGTCTAGTGGCTGGCTGAGGATTCAAGAAAGT 1792
QY 1629 TAATTCAGTGTATGTTCAGGAACTTCAATGTCTTGTCCACATGTTCAGGAAATCACTGC 1688
DB 1793 AAATTCAGTGTCTTATCAGGAACTTCAATGGCTGTCTCATGTAGTGGCAATGCTGTC 1852
QY 1689 TCTTATCCGGTCTCATACCCGAACTGCTCTCCAGCTGCAATCAAAATCCGCTATTGATGAC 1748
DB 1853 ACTACTCAATTCATTTCTTAAATGCTCACAGCTGCAATCAAAATCCGCTTAATGAC 1912
QY 1749 AACACGGGATTTGTACGATCGTCAAGGGAAGCGATAAAGGATGGTAAACAAACCGCCGG 1808
DB 1913 AACTGCAGACACAAACACCAAGGAAACCAATCATGATGTTGACACACAGGCTGG 1972
QY 1809 TGTGTTTGGATTTGGAGCAGGCAATGTAATCCGCAAAAGCGGATAAACCGGGAATGGT 1868
DB 1973 ACTTTCCGCAATAGGAGCTGGACATGTAATCTCTGGAAGATCCGATGATCCCGGATTTGAT 2032
QY 1869 TTACAAACATTCACAGTGGATTACATAAGCTTACTCTGCACTCTTGGATTTCACAGATC 1928
DB 2033 ATATGACATTAATGCAATGACTATATACCTCACCTTTGCACTATTGTTTACAAAACTC 2092

QY 1929 AGATATTTAGCAATCACTCATAGAACGCTGAGCTCAATGGAATATTCGGAAAAACCC 1988
DB 2093 TGAATCCTCAGCATTAATCAAGAAATGTTAGCTCCAGCAGTTTACAGAAAAACAG 2152
QY 1989 GGGTTTATGCTCAATACCCGTCGATAGCCGATGATTTCAAAAGTGGAAGCACTACGGA 2048
DB 2153 GGGTTTATGCTCAATACCCCTCTATTTCCGTAATCTTTAAGGCAGGAAAAACGAGAA 2212
QY 2049 GATGATCACAAAGCGGTGCTACTAAAGCTGGAGTCTCACTGATATCTGATGATGCT 2108
DB 2213 AATGATCACAAAGAGAGTGCATAATGTGGGAGTCTCAATTAATCTACTCAGTTGAAT 2272
QY 2109 CAAGGCTCCAGAGGGGATCAAGATTTATGTCATCTCAAGAGACTTGTGTTCAAAACAGT 2168
DB 2273 TGTGGCACCAGAGAGGTTAAAGTGAGAGTTAAACCGCAGCTGCTGTTTAAACATGT 2332
QY 2169 GGATCAGACCTGAGCTATAGATGATGTTGTTGATTTGAAGAGAAAAACAGAGAGGAA 2228
DB 2333 TAATCAAGTTTAAAGTTACAGAGTTGGTTTATATCAAGGAAGAGA--ATTGGGACTCA 2389
QY 2229 GGTGGCTAGTTTGCACAGGGGAGTTGACTGGGTCAACTCTCATATCTGATGACGG 2288
DB 2390 AAGGAGAGCTTTGCAAGAGGACAATGATGTGGATCAACTCCAGAGATAAATACAGAA 2449
QY 2289 AGTTAGAAGPCCAAATCTCTGAACCTTGAAGACTAA 2324
DB 2450 AGTTAGAAGTCTTATTTAGTTCCATGGGCATCAA 2485

RESULT 4

AAZ93814
ID AAZ93814 standard; DNA; 3140 BP.

XX AAZ93814;

XX 16-AUG-2000 (first entry)

XX Subtilase (St_P2Sca) of Solanum tuberosum.

XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
KW crop protection; feed; foodstuffs; ss.

XX Solanum tuberosum.

XX Key Location/Qualifiers
FT CDS 1..2301
FT /*tag= a
FT /product= Subtilase St_P2Sca

XX W0200022144-A2.

XX 20-APR-2000.

XX 12-OCT-1999; 99WO-EP07633.

XX 12-OCT-1998; 98EP-0119244.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

XX Berger D, Altmann T;

XX WPI; 2000-317995/27.

XX P-PSDB; AAY83304.

XX Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
PT lower water consumption and enhanced diseased resistance

XX Claim 1; Page 85-88; 101pp; English.

CC Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO₂ uptake into and H₂O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This subfamily of
CC Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
CC (See AAZ93808).

XX SQ Sequence 3140 BP; 939 A; 627 C; 690 G; 882 T; 2 other;

Query Match 43.9%; Score 1020.2; DB 21; Length 3140;
Best Local Similarity 66.4%; Pred. No. 1.4e-307;
Matches 1530; Conservative 1; Mismatches 754; Indels 18; Gaps 4;

QY 14 CTCTCTCTCTCGCATATCTTTCTTCTATTTTGTCTCTCTCGTCAGAGATCCTCGAGA 73
DB 2 CTCATTTATCTCTTCTATGCTTTTACTATGTTTGTTCATACAGCTCAGATT 61
QY 74 AGCAGACTTACATGTTTCAAGCTTCATCTTAATAGCGAAACCGCTAAACCTTGCCTCA 133
DB 62 TGCAAACTTACATAGTTCACTTACATCCACATCGAGCAACAGACCCCTTTAGCTCTA 121
QY 134 AGTTGATTGGCATCTTTCTTTCTCCAGAGCGGTTTGTAGTGTGAGAGAGAGG 193
DB 122 AACTACAATGCGACCTTCTTCTCTTGCAGAGCAGTTCTCTCT-----GGAGAAC 172
QY 194 AAGAGCCTCTCTCGACTTCTCTACTCTCTATGGCTCTCGATTGAGGATTTGCTGCTC 253
DB 173 AAGACTGCTTCTCTGCTTTTGTACTCTTACATCTCGATGGAAGGTTTTCAGCTC 232
QY 254 AGTTGACTGAATCAGAAAGCGAGATCTAGATATTCACCTCAAGTTGTTGCACTGAGAC 313
DB 233 GACTCACTGAAGATGAGTTGAGTTGTTAAGGGAATCTAATGATGTTGTCATACGTG 292
QY 314 CTGACCATGTTCTTCAAGTTCAAAACCATCTACTCTTACAAAGTTCTTGGGACTCGAGGT 373
DB 293 CTGAGAGAGGCTTGAATTCAGACTACTTATTTTCAAGTTCTTGGGATT---AAGTC 349
QY 374 TTGGAACTCCGGTGTATGGTCTAAATCTCGTTTGGTCAAGGCAATATCGCGGTGC 433
DB 350 CAACGAGAGAGAGGCTTGGTTGAAGTCTGGATTGGTCAGGGGCGATCAITGGAGTGT 409
QY 434 TTGATCTGAGTTTGGCTTGAAGTCTTAGCTTTGACGATACCGGAATGCCCTTCGATTTC 493
DB 410 TGGTACTGAGTTTGGCCAGAAAGTCCAAAGTTTTCATGATCATGSGATGCCACCTGCTC 469
QY 494 CACGAAATGGAAGGGATTTCGCAAGAGGAGAAAGTTTCAGTTCTTCGAGCTGTAACC 553
DB 470 CACAGAGTGGAGGGTGTCTGCCAAGGAGGACAGGATTTAATCTTCTAGTTGTAATC 529
QY 554 GGAAGCTAATCGTGTAGATTCTTTCATCAGAGGACACCGTGTCTGCTAATTCACAGAGG 613
DB 530 GCAAGCTTATTTGTCGAAGGTTTTCAGAAAGAGACATCTGTGGC---TTCAATGACAT 586
QY 614 AATCACCACCATGCTCTGATACATTTCCGCAAGAGATTCACGCGGACACGGGACTC 673
DB 587 CATCACCAGATGCAAGTGGAGGAATATGTGTGCCACCGGATTTCCCATGCCATGGTACAC 646
QY 674 ACACCGCTCAACAGTTGGTGGATCCTCTGTTTCGATGGCGAATGTTTCTTGGCAATGGAG 733
DB 647 ATACAGCATCACTGCTGGAGAGCTGCACTTCCATTTGGCTGTGTGCTCGGAATGGAG 706
QY 734 CTGTTGTGGCTGTGGGATGGCTCTCGAGAGCTCACATTCGAGTCTATATAAGTCTGTGGT 793
DB 707 CAGGGAGGCTCGAGGAGTGGCCCGGTCGCCACATTCGCAATATATAAGTATGCTGTT 766
QY 794 TCAATGTTGTTACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCATTCAGATTAAG 853

Db 767 TCAGTGGTTGTTACAGCTCTGATATATCTTCAGCAATGGATGTGGCCATCAGAGATGGAG 826
Qy 854 TCGATGTTCTTTTCGCTTTCCCTTGGCGGTTCCCTATTCCTTTGTATGATGACACAAATCG 913
Db 827 TAGACATATTGTCACTCTCACTTGGTGGCTCCCTATTCCACTTTTATGATGATCTATTG 886
Qy 914 CCATTGGAACATTCAGAGCCATGGAACGCGGTATATCTGTAACTGTGTCAGCTGGTAAACA 973
Db 887 CCATTGGAAGTTTCCAGGCCATGAGCATGGAAATTCAGTTATATGTCTGTCAGGGAATA 946
Qy 974 ACGGTCCAAATCGAAAGCTCTGTTGCAACACACAGCTCCTCGGTCTCAACCATTTGGCGCAG 1033
Db 947 ATGACCAATCCAAAGTTCACTAGCCCAACCGTGTCTCTGATGATGCACTATTGGTGTCTA 1006
Qy 1034 GCACGTTGATCGAAGATTTCCCGCTGTGTCTAGATTTAGCCCAACGGAAGCTTCTCTATG 1093
Db 1007 GCACATTGACAGGAGATTTCCAGCGTCAGTTAGGCAACGGAAGTTCTCTCTACG 1066
Qy 1094 GAGAGTCATTGTATCCGGGAAGGTATAAAGAAATGCCGGGAGAGAGTTTGAGGTGATTT 1153
Db 1067 GAGAATCCTTTGTACCTCGGGAAGAAATTCCTAGCTCTCAGAAGAATCTTGAGATCGTTT 1126
Qy 1154 ACGTTCACAGAGGAGATTAAGGAGTCACTCTGTTTGTAGAGGTCACCTTCCAAGAGAG 1213
Db 1127 ATGTAAGGATAGGACAGAGGAAGTGAATTTGCTTGAGAGATCGCTATCAAAAGCAC 1186
Qy 1214 AAATCCGAGGCAAAATCGTGAATTTGTATCGCGAGTCAATGGAAGATCGGAGAAAGGAG 1273
Db 1187 AAGTCCGAGGGAATCGTTGTGTGTATAGGGAGTCAATGGAAGGAGGAGGAGGAGG 1246
Qy 1274 AAGCGGTTAAGAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCAAG 1333
Db 1247 AGGTTGTGAAGGAGGAGGAGTGGTGTCTCCATGATCTTAGCAATACAGCAATAAATATG 1306
Qy 1334 AAGAGATTTATGAGGTTCACTCTTACAGCTACATTTGATGTTTACATGAGTCAAG 1393
Db 1307 AGGAAGATTCATTTGATGTCATGTCTCCAGCAACGTTGATTTGCTTCGATGAATCAA 1366
Qy 1394 TCCTTCGAGGCTTATGTTAAATGCCAGGTTGAACCAAGGCGCGGATATTTTGGTG 1453
Db 1367 TTCATTACAAACTCTGAACTCAACAAAAGACCAACAGCTCGATTCATATTTGGAG 1426
Qy 1454 GTACGTTGATGGAGGTCACAGACCGGAGGTGGCTCAGTTTTCAGCTTCGAGGACCGA 1513
Db 1427 GAACGGTAATAGGAAGCTTAGAGCACTGCAGTAGCTCAGTTTTCGTCAGGGGGCCAA 1486
Qy 1514 GTTTAGCCATCTCTCGATCTAAACCGGATATGATGCTCCGGAGTCAATATCAATG 1573
Db 1487 GCTATACGTGATCCTTCAATTTCTAAACCTGATTTGATGCTCCAGGGTAAACATAATTG 1546
Qy 1574 CGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTTCAAGAGAGTTAACT 1633
Db 1547 CCGCTTGGCCACAAAATCTAGGCCCCCAGTGTCTTCCGAAGATTCACGAAGAGTAAAT 1606
Qy 1634 TCAGTGAATGTAGGAACCTCAATGCTTGTCCACATGTTAGCGGAATCACTGTCCTTA 1693
Db 1607 TCAGTGTATGTAGGGACCTCAATGCAATGCTCTCTATGATTTCAAGAGAGTTAACT 1666
Qy 1694 TCCGCTCTGCATACCCGAATGCTCTCCAGCTGCAATCAATCCGATTTGATGACAAACAG 1753
Db 1667 TCCATTGAGTCTATCTTAATGGACTCCAGCAAGTAAGATCCGCAATTAATGACCACTG 1726
Qy 1754 CGGATTTGATGATCGTCAAGGGAACCGGATAAGGATGGTAAACAAACGCGGTGTGT 1813
Db 1727 CAGATACAGCTGATCATATGGGAAACCAATCATGTGATGAGATGACACCAAGCTAAACTTT 1786
Qy 1814 TTGCGATTGAGGAGGCGCATGTAATCCGCAAGGCGATAAACCCGGGATGGTTTACA 1873
Db 1787 TTGAGCTGAGCTGGACAGCTGAACTCCGGAAGAGCCATCGATCTCTGGATTTGATATG 1846
Qy 1874 ACATTCACAGCTGGATTACATAACTTACCTCTGCACTCTTGGATTCACAAAGATCAGATA 1933
Db 1847 ACATCCAGTTGATGATATATCACTCATCTTTGCACTATCGGATACAGAAATCTGAG 1906

RESULT 5

AAZ93812

ID AAZ93812 standard; DNA; 3865 BP.

XX AAZ93812;

XX AAZ93812;

DT 16-AUG-2000 (first entry)

XX Subtilase (Pigesp) of Solanum tuberosum.

KW SDD; serine protease; subtilisin; transgenic plants; dry weight;

KW stomata; sugar; water; protein; CO 2; H 2O; CO2; H2O;

KW crop protection; feed; foodstuffs; ss.

XX Solanum tuberosum.

FH Key Location/Qualifiers

FT exon 3..551

FT /tag= a

FT /label= Exon 1

FT 552..986

FT /tag= b

FT /label= Intron 1

FT 967..1654

FT /tag= c

FT /label= Exon 2

FT 1655..1737

FT /tag= d

FT /label= Intron 2

FT 1738..2222

FT /tag= e

FT /label= Exon 3

FT 2223..2485

FT /tag= f

FT /label= Intron 3

FT 2486..3252

FT /tag= g

FT /label= Exon 4

XX WO200022144-A2.

XX 20-APR-2000.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PF 12-OCT-1999; 99WO-EP07633.
XX
PR 12-OCT-1998; 98EP-0119244.
XX
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
FA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
XX Berger D, Altmann T;
XX
XX WPI; 2000-317995/27.
DR P-PSDB; AAY83302.
XX
XX Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
PT lower water consumption and enhanced diseased resistance
XX
XX Claim 1; Page 72-76; 101pp; English.
XX
XX Sequences encoding SDP1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO₂ uptake into and H₂O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This subfamily of
CC Solanum tuberosum is a homologue of SDP1 of Arabidopsis thaliana
CC (See AA293808).
XX
XX Sequence 3865 BP; 1202 A; 707 C; 715 G; 1237 T; 4 other;
SQ
Query Match 25.9%; Score 602; DB 21; Length 3865;
Best Local Similarity 58.5%; Pred. No. 9.5e-177;
Matches 1344; Conservative 0; Mismatches 606; Indels 349; Gaps 3;
372 TTTTGGAACTCGGTGTATGCTAAATCTCGTTTGTCTCAAGGACAAATATCGGGT 431
951 TTTTCTTCTAATAGGTACTTGGTTAAAGTCTGGAATTTGTGAGGGCGCATCTGGAGT 1010
432 GCTTGATCTGAGTTGGCTCAAGTCTAGCTTTGACGATACCGGAATCCCTCGAT 491
1011 TCTTGATCTGGAATTTGGCCAGAAAGTCAAGTTTGTGATCATGGAATCTCCTAT 1070
492 TCCACGAAATGGAAGGATTTGCCAAGAAGGAGAAAGTTTCAGTCTTCGAGCTGTAA 551
1071 TCCAAAGAAATGGAAGGTTCTGCCAAGAAGGAAAACTTCAATCTTCAAGTTGCAA 1130
552 CGGAAGCTAATCGGTCTAGATTTCTTCATCAGAGACACCGTGTCTAATTCACAGA 611
1131 TCGCAAGCTTATTTGGTCAAGTTTTCAGATAGGACACATGATGGCATCAAGACATC 1190
612 GGAATCACCACAAATCGCTGTGAATACATTTCCGCAAGAGATTCACGGGACAGGAC 671
1191 AAATCAATAGATTTTATGGAGGATATGATCACCCTCGAGATTTCAAGGGCCATGGTAC 1250
672 TCACACCGCTCAACAGTTGGTGGATCTCTGTTTCGATGGGGAATGTTCTTGGCAATGG 731
1251 ACATACAGATCTACTGACGGGGAGCTCCGTTCCAAATGGGAGTGTCTTGGAAATGG 1310
732 AGCTGGTGTGGCTCGTGGGATGGCTCTGGAGTCAATGCAATGATGATTAAGTCTGTTG 791
1311 AGCAGGAGAGGCTCGAGGATGGCCCTGGTGTCTATATCGCGATATACAAAGTTTGTG 1370
792 GTTCAATGGTTGTACAGCTCTACATTTAGCAGCTATAGATGATGAGCAATTCAGATAA 851
1371 GTCTAGTGGTGTATAGTTCTCATATCTACTTGCAGCAATGGATGATGATGATGATG 1430
852 AGTCGATGTTCTTTCGTTTCCCTTGGCGTTTCCCTATTCTCTTCTATGATGACACAT 911
1431 AGTAGACATATTTCTCTTTCAATTTGGTGTGTTTCCCTGTTCCACTTTATGAGGATCTAT 1490

QY 912 CGCAATTTGGAACATTCGAGCCATGGAACGCGGTATATCTGTAAATCTGTGACGCTGGTAA 971
DB 1491 TGCTATTTGGCAGTTTTCGAGCTATGGAACGCTGAAATTCAGTTATATGCTGCTCAGGAAA 1550
QY 972 CAACGGTCCAAATCGAAAGCTCTGTTGCAACACAGCTCCTTGGTCTCACCAATTCGGCG 1031
DB 1551 TAATGGTCCAATTTCAAGTTTCAGTAGCAATGAGGCTCCTTGGATTCACACTATGGTGC 1610
QY 1032 AGGACGCTTGTATCGAAGATTTCCCGCTGT----- 1061
DB 1611 TAGCACACTTGACAGGAAATTTCCAGCAATTAATTCAGTAGGTATGTGCACATTTTGTTC 1670
QY 1062 -----GGTCAGA 1068
DB 1671 TTHAATGATATTTCCGGTGTTCAGCCTAAATATGTCTCCTCATTCATATTTTCCA 1730
QY 1069 TTAGCCAAACGAAAGCTTCTATGAGAGTCAATTTGATCCGGGAAAAGGTATAAAGAT 1128
DB 1731 ACAGTTAATGGCAAGTATGTATGGAGAAATCTTGTACCCGGGCAACAAGTTCATAAT 1790
QY 1129 GCGGGAGAGAGTTGAGGTGATTACGTCACAGGAGGAGATAAAGAACTGAGTCTGT 1188
DB 1791 TCTCAGAAAGTTCTTGAGATGTTTATCTCAATGACGGTGATTAATGAAAGTGAATTTGC 1850
QY 1189 TTGAGAGGGTCACTTCCAGAGAGAAATCCGAGGCAAAATGTTGATTCGCGGA 1248
DB 1851 TTAAGAGGGTCTCTGCCAAGAGCTAAGTCCATGGAAATTCGTTGATGATCGTGA 1910
QY 1249 GTCAATCGAAGATCGGAGAAAGAGAGCGGTTTAAAGAACTGAGGAGTTGCAATGATC 1308
DB 1911 GTTAATGGAGAGCAGAGAAAGGTCAAGTTGTTTAAAGAAATCAGGTGTTGTCATGATC 1970
QY 1309 TTAGCCAAATACAGAGATCAACCAAGAGAGATTTCTATTGACGTTTCATCTTACCAGCT 1368
DB 1971 CTAGCAAAATACAGAGTAAATATGAGGAGAAAGTCTGTGACGATCATGTCCTACCTGCA 2030
QY 1369 ACATTGATTTGGTTACACTGAGTCAGTCCCTTCTCAAGGCTTATGTTTAAAGCAAGGTTG 1428
DB 2031 ACATTGATTTGGTTGACGAATCAATTCAGTTCCAAAGCTATATGAACTCAACGCAAAA 2090
QY 1429 CCAAGGCGGGGATTAATTTTGGTGTACGTTGAGTGGGAGGTCAAGACACCGGAGGTG 1488
DB 2091 CCAACAGCTCGAATCATATTTGGAGGAAACAGTTTATAGGAAATCTAGTGCACCTGTGTA 2150
QY 1489 GCTCAGTTTTCAGTCTCAGACCGAGTTAGCCAAATCCTTCGATATAAACCAGGATATG 1548
DB 2151 GCACAAATTTCTTCAAGGGGTCCAAGTTTACTGATCTTCAATTTCTCAAAACCTGATG 2210
QY 1549 ATTGCTCCG----- 1557
DB 2211 ATTGCTCCAGTCAAGTTTATTTGAACCCCAATTTATTTTAGATCATAGTAGCAAT 2270
QY 1558 ----- 1557
DB 2271 GTGACCAACAGTTAGGATTCGAGCGTGGAAATCTATCATTTGATGCTGAATCAGGTA 2330
QY 1558 ----- 1557
DB 2331 GACTGCTTACATCACACCTTCCACAGACCATGATGAAACCGGATGCTCTTTTATAT 2390
QY 1558 ----- 1557
DB 2391 GCATGTGAAAAAACTTTAATAAATAGTGAATGTTATGTTTGAACCTATATCTTCGTA 2450
QY 1558 -----GGAGTCAATATCATTCGGCTTGGGCTC 1585
DB 2451 TAATCAGAAATCTGAAATCCGCCCTCTCTCAGGTGTCAACATAATGCTGCTTGGCGC 2510
QY 1586 AAAATCTAGGACCAACCGGACTTCTTTATGATTAAGTCAAGAGAGTAACTTCACGTGAATG 1645
DB 2511 AAATCTAGTCTTAGTGGCTCTGCTGAGGATTCAGAGAGTAACTTCACGTGCTTAT 2570
QY 1646 CAGGAATCTCAATGCTTGTTCACATGTTAGCGGAATCACTGCTCTTATCCGGTCTGCAT 1705

```
Db 2571 CAGGAACCTTCAATGGCTGTCTCATGTAGTGSCATTGTGCACACTCCATTCATTC 2630
Qy 1706 ACCCGAAGTGTCTCCAGCTGCAATCAATCCGCAATGATGACAAAGCGGATTTGTACG 1765
Db 2631 ATCCCTAAATGGTCCAGCTGCAATCAATCCGCGCTAATGACAACTGCGACACACAA 2690
Qy 1766 ATCGTCAAGGGAAGCGGATAAGGATGTAACCAACCGCGGTGTGTTCGATTTGGAG 1825
Db 2691 ACCACCAAGGNAACCAATCATGTGATGTGACACAGAGCTGGACTTTTCGCCATAGGAG 2750
Qy 1826 CAGGCGATGTGAATCCGCAAAAGCGGATAAACCCGGGATTTGGTTTACCAATTCACACCGAG 1885
Db 2751 CTGGACATGTAAATCCCTGGAGATCCGATGATCCCGGATTTGATATGACATTAATGCCAA 2810
Qy 1886 TGGATTACATACTTACTCTCTGCACTCTTGCACTTGGATTCAAGATCAGATATTTAGCAATCA 1945
Db 2811 ATGACTATATCACTCACTTCTGCACTTCTGCACTTCTGCACTTCTGCACTTCTGCACTT 2870
Qy 1946 CTCATAGAACGTCGAGTGCATGCAATGGAATATGCGGAAACCCGGGTTTGTCTCAAT 2005
Db 2871 CTCACAAAGATGTGATGTCGACGACGTTTACAGAAACACGGGGTTTGTCTCAAT 2930
Qy 2006 ACCCGTCGATAGCGGTGATTTTCAACGTCGCAAGACTACGGAGATGATCACAAGCGGTG 2065
Db 2931 ACCCTCTATTTCCGTAACTTTAAGCAGGAAACACGAGAAATGATCACAAGGAGAG 2990
Qy 2066 TCACCTAAGTGGGAGTCCCTAATCTGATATCTGATGATGTCAGGATGTCAGAGGGGA 2125
Db 2991 TGACAAATGTGGGAGTCCCTAATTTCAATCTACTGATGTAATTTGGGACACGAGAGAG 3050
Qy 2126 TCAAGTATTTCAATCTTCAAGAGACTTGTGTTCAACACGTCGATGATGAGCTGAGCT 2185
Db 3051 TTAAGTGAAGATTAACCCGCGACGTCGTGTTTAAACATGTTAATCAAGTTTAAGTT 3110
Qy 2186 ATAGATGATGTTGTTGATTTCAAGAGAAACAGAGAGGAGGAGGTCGCTTGTTCAC 2245
Db 3111 ACHAGTTTGGTTTATATCAAGAGAGA---ATTGGGACTCAAGGAGAGGAGGTCGCTTGTTCAC 3167
Qy 2246 AAGGCGAGTTGACTTGGGTCACTCTCAATCTGATGATGATGATGATGATGATGATGATGAT 2305
Db 3168 AAGGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3227
Qy 2306 CTGTAACCTTGAAGACTAA 2324
Db 3228 CAGTTGATGGCATCAAA 3246

RESULT 6
AAC43317
ID AAC43317 standard; DNA; 2265 BP.
XX AC AAC43317;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 38925.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN BP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000BP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
```

```
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137582.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
```


PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.	QY	380	ACTCGGTGTATGTCTAAATCTCGGTTTGGTCAAGGCACAATATATCGGCGTCTTGATA	439
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.	PR	08-OCT-1999;	99US-0158232.	
PR	16-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.	PR	12-OCT-1999;	99US-0158369.	
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.	PR	13-OCT-1999;	99US-0159293.	
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.	PR	13-OCT-1999;	99US-0159294.	
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.	PR	13-OCT-1999;	99US-0159295.	
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.	PR	14-OCT-1999;	99US-0159330.	
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.	PR	14-OCT-1999;	99US-0159331.	
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.	PR	14-OCT-1999;	99US-0159637.	
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.	PR	14-OCT-1999;	99US-0159638.	
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.	PR	18-OCT-1999;	99US-0159584.	
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.	PR	21-OCT-1999;	99US-0160741.	
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.	PR	21-OCT-1999;	99US-0160767.	
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.	PR	21-OCT-1999;	99US-0160768.	
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.	PR	21-OCT-1999;	99US-0160770.	
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160814.	PR	21-OCT-1999;	99US-0160814.	
PR	23-JUL-1999;	99US-0145145.	PR	21-OCT-1999;	99US-0160815.	PR	21-OCT-1999;	99US-0160815.	
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160980.	PR	22-OCT-1999;	99US-0160980.	
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.	PR	22-OCT-1999;	99US-0160981.	
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.	PR	25-OCT-1999;	99US-0161404.	
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.	PR	25-OCT-1999;	99US-0161405.	
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.	PR	25-OCT-1999;	99US-0161406.	
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.	PR	26-OCT-1999;	99US-0161359.	
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.	PR	26-OCT-1999;	99US-0161360.	
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.	PR	26-OCT-1999;	99US-0161361.	
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.	PR	28-OCT-1999;	99US-0161920.	
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161922.	PR	28-OCT-1999;	99US-0161922.	
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.	PR	28-OCT-1999;	99US-0161993.	
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.	PR	29-OCT-1999;	99US-0162142.	
PR	05-AUG-1999;	99US-0147192.							
PR	05-AUG-1999;	99US-0147260.							
PR	06-AUG-1999;	99US-0147303.							
PR	06-AUG-1999;	99US-0147416.							
PR	08-AUG-1999;	99US-0147493.							
PR	09-AUG-1999;	99US-0147935.							
PR	10-AUG-1999;	99US-0148171.							
PR	11-AUG-1999;	99US-0148319.							
PR	12-AUG-1999;	99US-0148341.							
PR	13-AUG-1999;	99US-0148565.							
PR	13-AUG-1999;	99US-014							

Db	878	GAGCGTTTGGCTATGAGAGAGGTGTTTTGTGTCGTGCTTGCCTGGTGAATAGTGCTC	937
Qy	980	CAATCGAAAGCTCTGTTGCAAAACACAGCTCCTTGGGTCTCAACATTGGCGCAGCAGCG	1039
Db	938	CTACTAGAGCTTCCTGTTGCCAATGTGCTCTTGGTTATGACTGTTGGTGTCTGGTACTT	997
Qy	1040	TTGATCGAAGATTTCCCGCTGTGTCAGATTAGCGAACAGAAAGCTTCTCTATGAGAGT	1099
Db	998	TAGATAGAGATTTTCCGGCTTTTTCGGAATCTCGGTAAACGGGAAACGACTTACCGGTGTT	1057
Qy	1100	CATTCTATCCGGAAAGGTATAAAGAAATGCCGGAGAGAGGTTGAGGTGATTTACGTCA	1159
Db	1058	CGCTGATAGCGGTGTAGGAATGGGGA-----CGAAGCCGCTTGAATGGTTTATAATA	1111
Qy	1160	CAGGAGGAGATAAAGGAAGTCAAGTTCTGTTTGAGAGGCTCACTTCCAGAGAAAGAAATCC	1219
Db	1112	AAGGAAATAGTAGTTCGAGTAATCTTTGTTTACCTGGTTCGCTTGATTTCGAGTATGTTTC	1171
Qy	1220	GAGCAAAATGGTGATTTTGGTGTGATCCGGAGTCAATGGAAGATCCGAGAAAGGAGAACGG	1279
Db	1172	GTGGGAAGATTGTTGTTGTGATAGAGGTGTAATGCTAGAGTTGAGAAAGGAGCTGTGG	1231
Qy	1280	TTAAGAGAGCTGGAGGAGTTTCAATGATCTTAGCCAATACAGAGATCAACCAAGAGAAG	1339
Db	1232	TTAGAGATGCTGGTGGTTTAGGAATGATAATGGCGAATACTGCTCGAGTGGAGAGGAGC	1291
Qy	1340	ATTCTATTGACGTTTCATCTCTTACACAGCTACATTTGATTGGTTACACTGAGTCAGTCTTC	1399
Db	1292	TTGTGGCGGATAGTCAITTTGTTCCCGGATCGCTGTAGGAAGAGACTGGTGATTTTAC	1351
Qy	1400	TGAAGGCTTATGTTAAATGCCACGGTGAACCAAGGCGCGATAATTTTTGGTGGTACGG	1459
Db	1352	TTAGGAGATATGTTAAAGTCAGATCTAAACCAACCGCTCTTCTGTTTTTAAAGGAACGG	1411
Qy	1460	TGATTGGAGGTACAGAGCACCGAGAGTGGCTCAGTTTTCAGTTCGAGGACCGAGTTTAG	1519
Db	1412	TTCTTTGACGTTTAAGCCGCTCTCTGTTGGTGTGCTTTAGCTCGAGAGGTCTTAATACGT	1471
Qy	1520	CCAATCCTTCGATACTAAAAACCGGATATGTTGCTCGGAGTCAATATCATTCGCGGCTT	1579
Db	1472	TTACTCTGAAATCTCGAGCGTGATGTTATTTGCTCTGGAGTTAATATTTTGGCTGGTT	1531
Qy	1580	GGCTCAAAATCTAGGACCAACCGGACTTCCTTATGATTCAAGAAGAGTTAACTTCACGT	1639
Db	1532	GGCTGACGCTATTGGTCTCTACTTGTCTTGCACAGGACTCTAGGAGGACTCAGTTCACA	1591
Qy	1640	TAATGTCAGGAATTTCAATGCTTTGTCACATGTTTAGCGGAATCATCTGCTTTATCCGGT	1699
Db	1592	TCATGTCAGGTACGTCAATGTCATGCCCCACATCATGAGTGGTTTAGCGGGTCTTTGAAAG	1651
Qy	1700	CTGCATACCCGAACTGTGCTCCAGCTCGAATCAATCCGATTTGATGACACAGCGGATT	1759
Db	1652	CAGCTCACCTCTAGTGGAGTCCGAGTGTATCAAAATCAGCTCTCATGACTACAGTTTAG	1711
Qy	1760	TGTACGATCGTCAAGGAAACGATAAAGATGGT-----AACAAACGACGGGTGTGT	1813
Db	1712	TTCTTGACAACACCAACGCTCCTCTCCATGATCTCGACACACAGCCTATCTAACCCAT	1771
Qy	1814	TTGCGATTGGAGCAGGGCATGTGAATCCGCAAAAGCGGATAAACCGGGATTGGTTTACA	1873
Db	1772	ATGCTCACGCTCGGGCCATGTAGATCCCAAAAGGCTCTCTCACGAGTCTTGTCTACG	1831
Qy	1874	ACATTCAACAGTGGATTACATAACTTACTCTGCACTCTTGGATTTCAGAGATCAGATA	1933
Db	1832	ACATCTCAACCGAGGAATACATCAGGTTTTTGTGCTCTCTAGACTACACAGTTCGATCACA	1891
Qy	1934	TTTTAGCAATCACTCATAGAACGTCAGCTGCAATGGAATATTGCGGAAAAAACCCGGTT	1993
Db	1892	TTGTTGGAT-----TGTGAGCGACCTAGCGTTAACTGCTCGAAGAAGATTCTCAGATC	1945
Qy	1994	TTAGTCTCAATTAACCGTTCGATAGCCGTGATTTTCAACGTCGCGAAGACTACGGAGATGA	2053

Db	1946	CTGCTCAGCTCAACTACCCAAAGTTTCTTCGGTTTCTTCGGTGTAAAGAGTTGTGCGGT	2000
Qy	2054	TCACAAAGCGTGTCACTAAGCTTGGGAGTCCCTAACTCGATATACTCAGTGAATGTCAAGG	2113
Db	2006	ACATCCGGAAGTAACAAATGTTGGTGCAGCAAGCTCGGTTTACAAAGTCACGGTTAAATG	2065
Qy	2114	CTCCAGAGGGGATCAAAAGTTATTGTCAATCTTAAGAGACTTGTGTTCAAAACAGTGTGATC	2173
Db	2066	GAGCTCCTAGTGTGGGAATCTCTGTTAAACCATCGAAACTTTTCGTTTAAAGCGTGGGAG	2125
Qy	2174	AGACGCTGAGCTATAGAGTATGGTTGTATTGAAGAAGAAAAACAGAGGAGGAAAGGTGG	2233
Db	2126	AGAAAGAGAGGTACACAGTCACGTTGTTGTTAGCAAGAAAGGAGTGAGTATGACGAAACAAG	2185
Qy	2234	CTAGCTTTTG 2242	
Db	2186	CTGAGTTTG 2194	
RESULT 7			
ABZ13260			
ID	ABZ13260	standard; DNA; 2265 BP.	
XX	AC	ABZ13260;	
XX	AC		
DT	21-JAN-2003	(first entry)	
XX			
DE	Arabidopsis thaliana	stress regulated gene SEQ ID NO 1065.	
XX			
KW	Arabidopsis thaliana;	plant; gene; stress; transgenic; ds.	
XX			
OS	Arabidopsis thaliana.		
XX			
PN	WO200216655-A2.		
XX			
PD	28-FEB-2002.		
XX			
PF	24-AUG-2001;	2001WO-US26685.	
XX			
PR	24-AUG-2000;	2000US-227866P.	
PR	26-JAN-2001;	2001US-264647P.	
PR	22-JUN-2001;	2001US-300111P.	
XX			
PA	(SCRI)	SCRIPPS RES INST.	
PA	(SYGN)	SYNGENTA PARTICIPATIONS AG.	
XX			
PI	Harper JF, Kreps J, Wang X, Zhu T;		
XX			
DR	WPI; 2002-304127/34.		
XX			
PT	Identifying a stress condition to which a plant cell has been exposed		
PT	and producing plants with increased tolerance to these abiotic stresses		
PT	-		
XX			
PS	Claim 144; SEQ ID NO 1065; 577bp + Sequence Listing; English.		
XX			
CC	The invention relates to identifying a stress condition to which a plant		
CC	cell has been exposed, comprising:		
CC	(a) contacting nucleic acid representative of expressed polynucleotides		
CC	in the plant cell with an array or probes representative of the plant		
CC	cell genome; and		
CC	(b) detecting a profile of expressed polynucleotides in the plant cell		
CC	characteristic of a stress response. The method is useful in the		
CC	production of transgenic plants, cells and seeds and in producing plants		
CC	with increased tolerance to abiotic stress. The present sequence is that		
CC	of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used		
CC	in methods of the invention.		
CC	Note: The sequence data for this patent is not represented in the printed		
CC	specification but is based on sequence information supplied to Derwent by		
CC	the European Patent Office.		
XX			
SQ	Sequence 2265 BP; 519 A; 510 C; 542 G; 694 T; 0 other;		

Query Match	20.7%	Score 481.8	DB 24	Length 2265
Best Local Similarity	55.3%	Pred. No. 2.7e-139		
Matches 1033	Conservative 0	Mismatches 812	Indels 24	Gaps 4
QY	380	ACTCCGGTGATGGTCTAAATCTCCGGTTTCGGTCAAGGCACAAATATATCGCGGTCTCTTGATA	439	
DB	344	AATTCGGTGTTACGATCTCCGGTCTCTCTCTAACGGCGTTATCATTCGGAGTTTATGATA	403	
QY	440	CTGCGAGTTTGGCCTGAAAGTCTTAGCTTTTGACGATACCCGGAATGCCCTTCGATTCACCGGA	499	
DB	404	CTGCGGTTTGGCCTGAATCTAGAAGCTTCGATGATCTGATATAGCTCTGAGATCCCTTCTTA	463	
QY	500	AATGGAAAGGGATTTGCCAAGAAGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGC	559	
DB	464	AATGGAAAGGAGAAATGTGAATCTGGTTCCGATTTCCGATTTCCAAAGTTGTGTAAACAAGAGC	523	
QY	560	TAATCCGTCGTAGATCTTCTCATCAGAGGACACCGTGTCCGCTAATTCACACAGAGGAATCAC	619	
DB	524	TTATCGAGCTGAAGCTTCTCCAAAGGATTTCAAATGGCTTCTCGTGGTGGTTTTCGA	583	
QY	620	CAAAACATGCCCTCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCG	679	
DB	584	GTAA-----GCGTGAATCTGTTTCTCTCTCGTGATGTTGACGGACATGGAACACATACTT	637	
QY	680	CCTCAACAGTTGGTGAATCCTCTGTTTCGATGCGGAATGTTCTTGGCAATFGAGAGCTGGTG	739	
DB	638	CAACTACCGCCGGGATCCGCGGTTAGAAACGCTAGCTTCCCTCGGTTACGCCCGCGGTA	697	
QY	740	TGCTCGTGGATGGCTCTCGGAGCTCACATTCGAGTCTATAAGTCTGTTGTTCAATG	799	
DB	698	CGGCCAGAGGATAGGCCACCTCGTCTCGTGTGCTACTTATAAGTTGTGAGGACTG	757	
QY	800	GTGTGTTACAGCTCTGACATTTAGCAGCTATAGATGTAGCGATTCAAGATAAAGTCCGATG	859	
DB	758	GTGTGTTTGGATCTGATATACTAGTCTGATGATGATCGAGCTATCTGTATGTTGTTGAUG	817	
QY	860	TTCTTTCCGTTTCCCTTGGCGGTTTCCCTAATCCTTTGTATGATGACACAATCGCCATTG	919	
DB	818	TGCTTTCCGTTATCTCTTGGTGGTGTCTGCTCCGCTATTATCGCGATACGATTTCCGATTG	877	
QY	920	GAACATTCGAGCCATGGACGGGTATATCTGTAACTCTGTCAGCTGGTTAAACAAACGGTC	979	
DB	878	GAGCGTTTTCGGGTATGGAGAGAGGTGTTTTGTGTTCTTCTGTCTGGTAAATAGTGGTCT	937	
QY	980	CAATCGAAAGCTCTGTGTCAAAACACAGCTCCTTTGGGTCTCAACATTGGCGCAGGACACGC	1039	
DB	938	CTACTAGAGCTTCTGTGCCAATGTGCTCTCTTGGGTATGATCTGTGTGCTGTGACTT	997	
QY	1040	TTGATCGAAGATTTCCCGCTGTGGTCAGATTAGCCAAACGGAAAGCTTCTCTATGGAGAGT	1099	
DB	998	TAGATAGAGATTTTCCGGCTTTTTCGCAATCTCCGTACCGGAAACGACTTACCCGGTGT	1057	
QY	1100	CATTGTATCCGGAAAGGATATAAGAAATCCGGGAGAGAGGTTGAGGTGATTTACGTCA	1159	
DB	1058	CGCTGTATACCGGCTGTAGGAATCGGGA-----CGAAGCCGCTTGAATTCGTTTATAATA	1111	
QY	1160	CAGGAGGAGATAAAGGAAGTGAGTTCGTGTTTGAGAGGTCACCTTCCAAAGAGAAGAAATCC	1219	
DB	1112	AAGGGAATAGTGTTCGAGTAACTTTGTTTAACTTGTTTCCGTTTGATTCGAGTATGTGTTT	1171	
QY	1220	GAGGCAAAATGGTGATTTGTGATCCGGAGTCAATGGAAGATCGGAAAGGAGAGACGG	1279	
DB	1172	GTGGGAAGATTGTTGTTTGTGATAGAGGTGTAAATGCTAGAGTTGAGAAAGGAGCTGTGG	1231	
QY	1280	TTAAGAGAGCTGGAGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCAAGAGAAG	1339	
DB	1232	TTAGAGATGCTGTGGTGTATAGGATGATAATAGCGCAATCTGCTGCCAGTGGAGAGGAGC	1291	
QY	1340	ATTCTATTAGCTTCATCTCTTACCACTCAATGATTGGTTACACTGAGTCAGTCCCTTC	1399	
DB	1292	TTGTGGCGGATAGTCATTTCGTTCCCGCGATTCGCTGTAGGAAAGAGACTGGGTGATTAC	1351	
QY	1400	TGAAGGCTTATGTTAATGCCACCGTGAACCAAGAGCGCGGATAATTTTTGGTGGTACGG	1459	

Db	1352	TTAGGAGTATGTTAAAGTTCAGATCTTAACCAACCGCTTCTCTGTTTTTAAAGGAACGG	1411
Qy	1460	TGATTGGGAGGTCACGAGCACCGAGGTGGCTCAGTTTTTTCAGTCCGAGGACCGAGTTTAC	1519
Db	1412	TTCTTGACGTTAAGCGCTCTCTGTGCTGCTTTTAGCTCGAGAGGTCCTAATACTG	1471
Qy	1520	CCAATCCTTCGATACTAAACCGGATATGATTGCTCCGGAGTCAATATCATTTGCCGCTT	1579
Db	1472	TTACTCCTGAAATCTTTGAAGCCTGATGTTATTGGTCTCGAGTTAATATTTTCGCTGGTT	1531
Qy	1580	GGCTCAAAATCTAGGACCAACCGGACTTCCTTATGATTCAAGAAGATTAACTTCACCTG	1639
Db	1532	GGTCTGACGCTATTTGGTCTCTATCTGCTTTGACAAGACTCTAGGAGGACTCAGTTCAACA	1591
Qy	1640	TAATGTGAGGAACCTTCAATGTCTTTGCCATGTTTAGCGGAATCACTGTCTTTATCCGGT	1699
Db	1592	TCATGTCAGTACGTCATATGTCATGCCACACATCAGTGGTTTAGCGGCTCTTTTGAAG	1651
Qy	1700	CTGATATCCGGAACCTGGTCTTCAGCTGCGAATCAAAATCCGATTTGATGACAAACGCGATT	1759
Db	1652	CAGCTCACCTCGASTGGAGTCCGAGTGCCTATCAAAATCAGCTCTCATGACTACAGCTTAAG	1711
Qy	1760	TGTACGATCTCAAGGGAAGCGATAAAGATGGT-----AACAAACCAAGCCGGTGTGT	1813
Db	1712	TTCTTGACAACAACAACCGTCTCTCCATGATCTCGACACAACAGCCTATCTAACCCAT	1771
Qy	1814	TTGCGATTGAGCAGGCGCATGTGAATCCGCAAAAGCGATATAAACCCGGGATTGGTTTACA	1873
Db	1772	ATGCTCACGCTCGGCCCATTGATATCCCAAAAGGCTCTCTCACCGAGTCTTGTCTACG	1831
Qy	1874	ACATTCAACGAGTGGATTACATACTTACCTCTGCACTCTTTGGATTTCACAGATCAGATA	1933
Db	1832	ACATCTCAACCGGAGGAATACATCAGGTTTTTTGTGCTCTCTAGACTACACAGTCGATCACA	1891
Qy	1934	TTTTAGCAATCACTCATAGAACGCTGAGCTGCAATGGAATATTGCGGAAAACCCCGGTT	1993
Db	1892	TTGTTGCGAT-----TGTGAAGCGACCTAGCGTTAACTGCTCGAAGAGTTCTCAGATC	1945
Qy	1994	TTAGTCTCAATTACCCTCGATAGCCGTGATTTTCAAACTGCGCAAGACTACCGGAGTGA	2053
Db	1946	CTGGTCAGCTCACTACCCCAAGTTTCTCGTTTTTGTGGGGTAAAGAGTTTGTGCGGT	2005
Qy	2054	TCACAAAGCGGTGCACTAACGTTGGGAGTCCCTAACTCGATATACTCAGTGAATGTCAAGG	2113
Db	2006	ACACTCGGGAAGTAAACAATGTTGGTGCACCAAGCTCGGTTTACAAAGTGACGGTTAAATG	2065
Qy	2114	CTCCAGAGGGGATCAAAAGTTATTGTCAATCTTAAGAGACTTGTGTTTCAACACGTGGATC	2173
Db	2066	GAGCTCCTAGTTCGGAATCTCTGTTAAACCATCGAAACTTCGTTTAAAGCGTGGGAG	2125
Qy	2174	AGAGCTGACCTATAGAGTATGTTTTGTATTGAAGAGAAAACACAGGAGGGAAGGTGG	2233
Db	2126	AGAAGAGAGGTACACAGTCACGTTTGTAGCAAGAAAGGAGTGAGTATCACGAAACAGG	2185
Qy	2234	CTAGCTTTG 2242	
Db	2186	CTGAGTTTG 2194	
RESULT 8			
AAC47609			
ID	AAC47609	standard; DNA; 2434 BP.	
AC	AAC47609;		
XX			
XX	18-OCT-2000	(first entry)	
XX	Arabidopsis thaliana	DNA fragment SEQ ID NO: 54454.	
XX	Hybridisation assay;	genetic mapping; gene expression control;	
KW	protein identification;	signal transduction pathway;	
KW	metabolic pathway;	promoter; termination sequence; ss.	

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence ss.

XX Arabidopsis thaliana.
OS EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149436.
PR 20-AUG-1999; 99US-0149732.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR	13-SEP-1999;	99US-0153758.	
PR	15-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	28-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 19.2%; Score 447.4; DB 21; Length 2434;			
Best Local Similarity 52.9%; Pred. No. 1.6e-128;			
Matches 1073; Conservative 0; Mismatches 926; Indels 30; Gaps 4;			
Qy	185	AAGAGAGGAGAGCCCTCTTCTCGACTTCTCTACTCTCTATGGCTCTCGATTGAGGAT	244
Db	300	AAGAAGAAGAAGTAATAACAACAGGATACTCTACACTTACCAGACTGCTTTCCAGGTT	359
Qy	245	TTGCTGCTCAGTTGACTGATCAGAGCCGAGATACAGATATTCACCTGAAGTTGTTG	304
Db	360	TAGCAGCTCAGCTTACTCAAGAAAGACAGAGAGGCTTGAGGAAGAAGTGTGTTGTAG	419
Qy	305	CAGTGACACCTGACCATGTTCTTCTCAGGTTCAAAACCACTTACTCTTACAAGTTCTTGGGAC	364
Db	420	CTGTGATACCTGAGACAGATACGAGCTTCACTACAGAGTCCAACTGTTCTTGGGT	479
Qy	365	TCGACGGTTTGGAAATCCGGTGTATGGTCTAAATCTCGGTTTGGTCAAGGACAAATTA	424
Db	480	TAGAAACACAAGAAGTGAGAGATTGGGCTCGAGAGAGTCAACCGCATGATGTGGTAG	539
Qy	425	TCGGCTGCTTGATCTAGTGGAGTTGGCTGAAAGTCTTAGCTTTGAGGATACCGGAATCC	484
Db	540	TTGGTGTGTTTAGACACTGGTATCTGGCTGAGAGTGAGACTTCAACGATACAGGTATGT	599
Qy	485	CTTCGATTCCACGGAAATGGAAGGGATTGGCAAGAAGAGAGAAAGTTTCAGTTCTCGA	544
Db	600	CCCCGTGTTCTGCTACTTTGAGAGAGAGCTTGTGAACTGGAAGAAAGATCTTGAAACGTA	659
Qy	545	GCTGTAAACCGAGAGCTAATCGGTGCTAGATTCTTCTACAGAGGACACCGGTGCTCGTAATT	604

Db	660	ACTGCAATAGAAAGATCGTTGGTGTAGAGTTTCTATAGAGG-----CTATGAG	710
Qy	605	CACCAGAGGAATCACCAAAACATGCTCGTAATACATTTCCGCAAGAGATTCAACGGGAC	664
Db	711	CTGCAACGGGAAGATCGATGAAGAGCTTGAATATTAAGTCAACGAGACAGAGATGGTC	770
Qy	665	ACGGGACTCACACCGCCTCAACAGTTGGTGGATCTCTGTTTCGATGGCAATGTTCTTG	724
Db	771	ACGGGACACACACTGCAGCTACTGTAGCTGGCTCACCTGTGTAAGGAGCTAATCTTTTG	830
Qy	725	GCAATGGAGCTGTGTGGCTCGTGGGATGCTCTCGGAGCTCACATTCAGTCTATAAAG	784
Db	831	GATTTGCTTATGGACAGCTCGAGGATGGCTCAAAAGGCTAGAGTTGCTGTATATAAG	890
Qy	785	TCTGTTGGTCAATGTTGTTACAGCTCTGACATCTTAGCAGCTATAGATGTAGCGATTC	844
Db	891	TCTGTTGGGTCGGAGGTTGTTTCAAGTTTCAACATTTTTCGGCTGTGATCAAGCTGTTG	950
Qy	845	AAGATAAGTCGATGTTCTTTCGCTTTCCCTTGGCGGTTTCCCTATTCCTTTGATGATG	904
Db	951	CTGATGGAGTTCAAGTACTCTCTATATCATTAGTGTGGGTTCTCTACTTATCTTAGAG	1010
Qy	905	ACAAATCGCATTTGGAACATTCGAGCCATGGAAACGGGTATATCTGTAATCTGTGCAG	964
Db	1011	ACAGTTTGTCTATAGCAACATTTGGAGCAATGGAGATGGAGTTTTCGTTTCGTGTTCTG	1070
Qy	965	CTGATAACAGGTCCTCAATCGAAGCTCTGTTCAACACAGCTCTCTGGTCTCAACCA	1024
Db	1071	CCGGTAATGGAGTCTCGATTCGATTCGATCTCACTAATGTTTCTCATGGATCACAAACAG	1130
Qy	1025	TTGSCGAGGACGCTTGTATCGAAGATTTCCCGCTGTGGTCAAGTTAGCCACGGAAGC	1084
Db	1131	TTGTTGCAAGTACTATGATAGAGATTTTCCAGCAAGTGAAGATAGAGACTATAGAA	1190
Qy	1085	TTCTCTATGGAGAGTCAATTCATCCGGAAGATTAAGAAATCCGGGAGAGAGGTTG	1144
Db	1191	CATTCAGAGAGTGTCTCTTCAAGGCGAAGACAGTTTTCCTTAAGATAAACAGATATC	1250
Qy	1145	AGGTGATTTACGT-----CACAGGAGAGATAAAGAAAGTGTCTCTGTTTGAGAG	1195
Db	1251	CTCTGGTTTACTTGGGAAGGAATGCAAGTAGTCTGATCAACCTCGTTCTGTCTAGATG	1310
Qy	1196	GCTCACTTCCAGAGAGAAATCCGAGGCAAAATGGTGTATTTGTGTCGGGAGTCAATG	1255
Db	1311	GAGCTTTGGATCGGCGCATGTAGCGGAAGATCGTGATATGGACCGCGGTGTACTC	1370
Qy	1256	GAGATCGGAAGAGAGAGCGGTTAAGAAAGCTGGAGGAGTTGCAATGATCTTAGCCA	1315
Db	1371	CACGTGTCAAAAGGTCAGGTTGTGAAGAGCTGTTGGAATGGGATGTTTAACTA	1430
Qy	1316	ATACAGAGATCAACCAAGAAAGATTCATTTGACGTTCTCTCTTACCAGCTACATGA	1375
Db	1431	ACACTGCAACAATGTTGGAAGAGCTTGTGCAGATTCCTCATATGCTTCCAGCTGTTGAG	1490
Qy	1376	TTGTTTACCTGAGTCACTCTCTGAGGCTTATGTTAATGTCACCGTGAACCAAGG	1435
Db	1491	TTGAGAGAAAGAGGTAATACTAATCAACAGTACGCGATGACGAGTAAAGAGCGACAG	1550
Qy	1436	CGCGGATAATTTTGGTGGTACGGTGTGGAAGCTCACGAGCACCGGAGTGGCTCAGT	1495
Db	1551	CGAGTTTAGAGATTTCTTGGAAACAAGATTTGATCAAACTTCCACAGTTGTAGACCGT	1610
Qy	1496	TTTCAGCTCGAGGACCGAGTTTAGCCAATCTTCGATACTATAAACCGGATATGATGCTC	1555
Db	1611	TCTCTTCAAGAGGACCAAAATTTCTGCTTTGGAGATCTTGAACCAAGAGCTTGTGGCTC	1670
Qy	1556	CGGAGTCAATATCATTTGGGCTTGGCTCAAAATCTTAGACCAACCGGAGTTCCTTATG	1615
Db	1671	CAGGAGTGAATATTTCTGACGCTTGGACTGGAGCATGGCACCATCGAGTTTATCATCTG	1730
Qy	1616	ATTCAAGAGAGTTTAACTTCACTGATGTCAGAACTTCAATGCTTCTTGTCCACATGTTA	1675

Db 1731 ATCCAAGAGGGTTAAGTTCATATATACTGTGGAACCTTCAATGTGTCATGTAA 1790
Qy 1676 GCGGAATCACTGCTCTTATCGGTTCTGCATACCGGAACCTGCTCCAGCTGCAATCAAT 1735
Db 1791 GTGGTGTAGCTGCTTTGATCAATCAAGGCATCCAGATTGGAGTCTCGCAGCAATCAAT 1850
Qy 1736 CCGCATTTGATGACAACAGCGGATTTGATCGTCAAGGGAAGGCATTAAGGATGGTA 1795
Db 1851 CAGCTCTCATGACAACCTGCTATGTTTCATGACAACATGTTTAAGCCTCTTACGGATGCAT 1910
Qy 1796 ACAAACAGCCGGT-----GTGTTTGGATTGGAGCAGGCGATGCAATCCGCAAAAG 1849
Db 1911 CAGGAGCAGCTCTTCAATCGCTTATGATCAAGTGCAGGACATATAGATCCTTTAAGAG 1970
Qy 1850 CGATAAACCCGGGATGGTTTAAACATTCACACAGTGGATTACATAAATCACTCTGCA 1909
Db 1971 CTACAGATCTGGTTGGTCTACGACATTCGACCTCAAGAGTATTTTGAATTCCTGCA 2030
Qy 1910 CTCCTTGATTCACAAGATCAGATATTTTGAATCACTCATTAAGAAGCTGAGTGCAATG 1969
Db 2031 CTCAGATTTAAGTCCATCAAGCTTAAGGTATTCACAAAACATTCAAAACAGAACTGCA 2090
Qy 1970 GAATATTGGGAAAACCCGGGTTTGTAGTCTCAATTAACCGTGCATAGCGGTGATTTTCA 2029
Db 2091 AACACACTCTTGCACGAATCCGGAACTTGAACCTACCCGGGATATCAGCTTTTCTCC 2150
Qy 2030 AACGTGCA-----AGACTACGGAGATGATCAAGCGGTGCATCAAGTTGGAGTC 2083
Db 2151 CAGAGAACACATGTTTAAGCTATGACATTTAAGAACAGTACCAATGTTGGTCTC 2210
Qy 2084 CTAACCTGATATCACTAGTGAATGTCAGGCTCCAGAGGGGATCAAGTTATTTGTCAATC 2143
Db 2211 ACATTTCAAGCTACAAGGTTTCTGTCGCCATTCAAAGCGCATCCGTAACGTGCCAG 2270
Qy 2144 CTAAGACACTTGTGTTCAACACAGTGGATCAGACGCTGACCTATAGAT 2192
Db 2271 CCAAAACACTCACTTCACTTCGAAGCACCAGAAGCTTTCTACACGGT 2319

RESULT 9
ID AAC46158 standard; DNA; 2637 BP.
XX AC AAC46158;
XX AC AAC46158;
DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49118.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX XX
PD 06-SEP-2000.
XX XX
PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.

Query Match 18.2%; Score 424.4; DB 21; Length 2637;

Db 975 TCATCTCTCTCCGTTGGAGCGTCGTGGTTCCTTATTACCTAGACGCTATCGCTATTG 1034
Qy 920 GAACATTCAGAGCAATGGAACCGGGTATATCTGTATCTGTGAGCTGTGTAAACAGGTC 979
Db 1035 GAGCTTTGGAGCTATTGACAGAGGATATTCGTCTCTGCTCCGCGGTAAACGAGGTC 1094
Qy 980 CAATCGAAAGCTCTGTTGCAACACAGCTCTCTGGGTCTCAACCATTTGGCGGAGGACGC 1039
Db 1095 CCGGTGCTTTAAGCGGTGACAAATGTTGCTCCGTGGATGACAAAGTCGGAGCTGGAACAA 1154
Qy 1040 TTGATCGAAGATTTCCGCTGTGGTCAGATTAGCCACGGAAGCTTCTCTATGGAGGT 1099
Db 1155 TCGATAGGGAATTTCCGCGCAATGTGAATCTCGGTAAACGGAAGATGATTTCTTGGTGTTA 1214
Qy 1100 CATTTATCCGGGAAGGTATAAA-----GAATGCCGGGAGAGAGGTTGAGGTGA 1150
Db 1215 GTGTATACGGTGGACCGGGTCTGATCCGGGTGCAATGTACCGCTGTTTACGGTGGTA 1274
Qy 1151 TTTACGTCACAGGAGGAGATAAGGAAGTGAATCTGTTTGGAGAGGTCACTTCCAAAG 1210
Db 1275 GTTATACGCGGCGATGGTTACTCATCGTCTCTGTGCTTGAAGGCTCGTTGGATCCGA 1334
Qy 1211 AAGMAATCCAGGCAAAATGGTATTTGATCGCGGAGTCAATGGAAGATCGGAGAAAG 1270
Db 1335 ATTTAGTGAAGGGAAGATCGTTCTTTGATAGAGGAATCAATCTAGAGCAACCAAG 1394
Qy 1271 GAGAGCGGTTAAGAGCTGGAGGAGTTCGAATGATCTTAGCAATACAGAGATCAACC 1330
Db 1395 GTGAGATCGTACGGAATAATGGAGGCTTGGGATGATTAAGCAATGGTGTGTTGAGC 1454
Qy 1331 AAGAAGAAGATCTATTAGCGTTTCATCTTTACAGCTACATTAATGGTTTACACTGAGT 1390
Db 1455 GTGAGGTTTAGTAGCTGATGCAACGTGTTACCGGCGACATCTGTTGGTCTTCGGAG 1514
Qy 1391 CAGTCCTTGAAGGCTTATGTTAAATGCCACGGTGAAA-----CCAA 1432
Db 1515 GAGATGAGATTCGTAGGTATATCTCTGAATCATCCAAATCTCGTTTCATCGAAACATCCAA 1574
Qy 1433 AGSCGCGGATAATTTTGGTGTGACGCTGATTTGGAGGTCACGAGCACCGAGGTGGCTC 1492
Db 1575 CGGCTACGATGTTTTCAAAGGTACCGGACTTGGGATTCGACCTGACCGGTTGTTGCAT 1634
Qy 1493 AGTTTTCAGCTCGAGGACCGAGTTTACGCAATCTTCGATACTAAACCGGATGATGTTG 1552
Db 1635 CTTTCTCTGCTGTGTCCTTAATCCAGAGCGCGGAGATCTTAAACCGATGTAATCG 1694
Qy 1553 CTCGGGAGTCAATATCATTTCCGCTTGGCTCAAAATCTAGACCAACCGGACTTCCTT 1612
Db 1695 CACCTGGTTTGAATATTCTAGCTGCTTGGCTGACCGGATTTGTTCCATCTGGTGTACTT 1754
Qy 1613 ATGATTCGAAGAGTTAACTTCACTGTATGTCAGGAATTTCAATGCTCTGTGTCACATG 1672
Db 1755 CTGATAACCGGAGAAGTGAATCAACATTTTATCAGGCACTTCGATGGCGTGCCCGCAGC 1814
Qy 1673 TTAGCGGAATCACTGCTTATCCGCTGCTCATACCGGACTGCTCCAGCTGCAATCA 1732
Db 1815 TGTCTGTTAGTCTGCTTGTGTTAAGCGGCTCATCCGATGAGTTCGCGCGCGGANTAA 1874
Qy 1733 AATCCGATTCATGACAAACAGCGATTTGACGATCGTCAAGGAAAGCATAAAGGATG 1792
Db 1875 GATCAGCATTTGATTAACACCGCTTACAGGTTGATACAGCGGTGAGCCGATGATGATG 1934
Qy 1793 -----GTAACAAACCGCGGCTGTTTGGATTTGGAGGAGGCGCATGTGAATCCGCAA 1846
Db 1935 AGTCCACTGCGCAATACATCTTCGGTTATGGATTATGTTCCGGGTCACTGTTCCACCAACCA 1994
Qy 1847 AGGCGATAACCCGGGATGTTTGAACATTAACAGCTGGATTAACATAACTTACCTCT 1906
Db 1995 AAGCTATGGATTCGGGATTAGTCTATGATATAACATCTTATGATTAATCAACTCTTGT 2054
Qy 1907 GCATCTTGGATTCAGAGATCAGATATTTTGAACAATCACTCAAGAAAGTGGCTGCA 1966
Db 2055 GTAAATCTTAACACTAGAACCAACATTTGTGACAAATAACCGGTGCGCAAGCGGACTGTG 2114

Qy 1967 ATGGAATATTGGGAAAAACCCGGGTTTTAGTCTCAATTACCGTCGATAGCCGTGATTT 2026
Db 2115 ACGTGTGAAGACGAGCGGACACGTCGGGAATTTAAACTACCGTCGTTCTCTGTCGTGT 2174
Qy 2027 TCAA-----ACGTGGGAAGACTACGGAGATGATCAAAAGCGGTGTCACCTAACGTTG 2077
Db 2175 TTCAGCAGTAGGAGAGAGTAATAATGTCGACACATTTTATTAGGACTGTGACAAATGTAG 2234
Qy 2078 GGAGTCCTTAATCGATATATCTCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTG 2137
Db 2235 CGGATTCGGATTCGTTTACGAGATTAAAGATTAGCGCGCTAGAGGACTACGCTGACGG 2294
Qy 2138 TCAATCTTAAGAGACTTGTGTTCAACACAGTGGATCAGACGCTGAGCTATAGATATGTT 2197
Db 2295 TTGAGCCGGAAGACTATCGTTTCAAGCGGTGGGGCAGAAACTGAGTTTTTGTGTTAGGG 2354
Qy 2198 TT 2199
Db 2355 TT 2356

RESULT 10
AAC42012
ID AAC42012 standard; DNA; 2638 BP.
XX
AC AAC42012;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 33960.
XX
DE Hybridisation assay; Genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151666.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match	17.9%;	Score 417;	DB 21;	Length 2638;
Best Local Similarity	51.5%;	Pred. No. 5.4e-119;		
Matches.1148; Conservative	0;	Mismatches 1030;	Indels 53;	Gaps 6;

Qy	13	CCTTTCCTTCTCTGCATATCTTTCTCTCTATTTTGTCTTCTTCTCGTCAGAGATCTCTGCAG	72
Db	137	CCTAACTTCTCTCTCAATGGCTTTCTTCTTACTTTCTTCTCTCTCTCTCTCTCTCTCTCTTCTT	136
Qy	73	AAGCAGACTTACATTTGTTCAGCTTCATCTCTTAATAGCGAAGCCGCTAAGACCTTTGCTCTCA	132
Db	197	CTCCTTCT	236
Qy	133	AAGTTTGATTGGCATCTTTCTTTCTCTCAAGAAGCGGTTTAGGTGTGTGAAGAAGAAGAG	132
Db	257	AAGCTAAACCTCAATCTTCCCCACTCATCTTCACTGGTACATCTCTCTCTCTCTCTCTCTCTCT	316
Qy	193	GAAGAGC--CTTCTTCTCGACTTCTTACTCTCTATGGCTCTGCGATTGAAGGATTTGCTG	250
Db	317	TCACATCTTCT	376
Qy	251	CTCAGTTGACGTGAATTCAGAACCGGAGATACGTGAGATATCCACTGAAGTTGTTTGCAGTGA	310
Db	377	CTGACTCACTTCAACAAGCGTAGTCAACTCTTTAGACCATCTCTCATGTCTCATCTCCGTTA	436
Qy	311	GACCTGACCATGTTCTTCAGGTTCAACCAACTTACTCTTACAGTTCTTGGGACTCGACG	370
Db	437	TCCCGGAACAAGTCCGTCATCTGCACACCACTCGTTCCTCCCTGAGTTTCTTGGACTCAGGT	496
Qy	371	GTTTTGGAACTCCGGTCTATGGTCTAAATCTCCGTTTGTGTCAGGACACAAATATCGGCG	430
Db	497	CCACCGCACAAAGCGGGTCTACTCGAAGAGTCTGATTTCCGGTCCGATCTAGTTATCCGGAG	556
Qy	431	TGCTTGATCTACGTGAGTTTGGCTCGAAAGTCTCTAGCTTTCAGATACCGGAATGCTTCGA	490
Db	557	TTATCGATCTAGTGTGTTTGGCCCGAAGACCTAGCTTTGATGACCGTGGTCTTGGTCTG	616
Qy	491	TTCCACCGAAATGGAAGGAGTTTCCCAAGAAGAGAAAGTTTCAGTTCTTCGAGCTGTA	550
Db	617	TTCCCATTAATGGAAGGCCAATGTATCGCTTCCCAAGATTTTCCGGAGTCTGCTTGTA	676
Qy	551	ACCGGAAGCTAATCGGTGCTAGATCTTTCATTCAGAGGACACCGTGTCTGCTTAATTCACAG	610
Db	677	ACCGTAAACTCGTCGGAGCTAGATTCTTCTCGCGTGGGTA-----TGAAGCAACCA	727
Qy	611	AGGAATTCACAAACATGCCTCGTGAATACATTTCCGCAAGAGATTCACCGGACACCGGA	670
Db	728	ACGGGAAATGAATGAACACGACTGAGTTTCGCTCTCCGCGGTGACTCCGATGGACATGGGA	787
Qy	671	CTCACACGCGCTCAACAGTTGGTGGATCCTCTGTCTTCGATGGCGAATGTTCTTGGCAATG	730
Db	788	CTCACACAGCTTCAATCTCCCGCGCGGTTAGCTTTTTCGGCGATCAACTCTCTGCGTACG	847
Qy	731	GAGCTGTGTGGCTCGTGGGATGGCTCTCGAGCTCACTTTCGAGTCTATAAAGTCTGTT	790
Db	848	CTCAGGTTGCTCGTCTGGGATGGCTCGAAGCTAGACTCGCTGCGGTACAAAGTCTGTT	907
Qy	791	GGTTCAATGGTTGTATACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCGATTCAAGATA	850
Db	908	GGAACTCCGGTGTGTACGACTCAGATATCTTTAGCGGCTTTCGACACCGCGGTTCCGACG	967
Qy	851	AAGTCGATGTTCTTTTCGCTTTCCCTTTGGCGGTTTCCCTATTCTCTTTGTATGATGACACAA	910
Db	968	GTGTGATGTCTATCTCTCTCTCCGTTGGAGCGGTGTGGTTTCCCTATTATACCTAGACGCTA	1027
Qy	911	TCGCCATTGGAAACATTCGAGCGCATGGAAACCGGTTATCTGTATCTGTGCACGCTGGTA	970
Db	1028	TCGCCATTAGGAGCTTTTGGAGCTATTGACAGAGAAATTCGTCTCTGCTTCCGCGCGGAA	1087

Qy	971	ACAA	CGCTCA	ATCGA	AGCTCTG	TGTC	CAACA	CACAGCTCCTTGGGTC	CTCAACCACTGGCG	1037	
Db	1088	ACGAGG	TC	CCGGTGC	TTTAAC	GGTG	CA	GAATGTGCTCGTGATG	ACAACAGCTCGGAG	1147	
Qy	1031	CAGC	ACGCTTG	ATCGA	AGATTTCC	CGCTGT	GTGC	AGATTAG	CCACGCGAAGCTCTCT	1090	
Db	1148	CTGGA	CAATCG	ATAGG	ATTTTCC	AGCCAA	TGTGA	AACTCGGTAA	ACGGAAGATGATTT	1207	
Qy	1091	ATGGA	GAGTCA	TTGTATCC	GGGAAA	AGGTATAA	-----	GAATG	CCGGGAGAGG	1141	
Db	1208	CTGGT	GTTAGT	GTATAC	GGTGG	ACCGGGT	CTGG	ATCCGGGT	CGAATCTACCCGCTGTTT	1267	
Qy	1142	TTGAG	TGATTTAC	GTCA	AGGAGAG	ATAAAG	GAAGTGA	TTCTGT	TTGAGAGGTCAC	1201	
Db	1328	ACG	TGTA	TTACTAG	GGGGGATGG	TTACTCAT	CGTCTCTGTGT	CTTTGA	AGGCTCGT	1327	
Qy	1202	TTCA	GAGAGAAATCC	GAGGCA	MAATGTG	TATTTGTG	ATCG	CGGAGTCAA	TGGAAGAT	1261	
Db	1328	TGGAT	CCGAATTTG	GTGAC	GGGAAAGATCG	TTCTTTGTG	ATAG	GGAATCA	TTCTTAAG	1387	
Qy	1262	CGGA	GAAGGAG	ACCGGTTAA	AGAACTGG	AGAGTGA	CAATGATCTT	TAGCCA	ATACAG	1321	
Db	1388	CAAC	AAAGTGAG	ATCGTAC	GGAATA	TGGAGCTTGGG	GATGATTA	TAGCGA	ATGGTG	1447	
Qy	1322	AGAT	CAACCA	GAAGA	AGATTTCA	TTCAGCGTTCA	TCTTTAC	AGCTAC	ATTTGATTTG	1391	
Db	1448	TGTT	TGACG	TGAAGGTTAG	TAGTCTG	ATTTGCC	ACGTGTTTAC	CGGCAC	ATCTGTCTGGTG	1507	
Qy	1382	ACAT	GAGTCAGT	CTTCTG	AGGCTTATG	TTAATGCC	ACCGTGAA	-----	-----	1428	
Db	1508	CTTCT	GAGAGAG	ATGAG	ATTCGT	AGGTATATCT	CTGAA	TATCA	TCAAAATCTGTT	1567	
Qy	1429	-----	CCAA	AGGCGCG	GATAATTTT	TGCTGGTAC	GGTGA	TGGAGGTC	CACGACACCGG	1453	
Db	1568	AAAT	CAACAGG	CTACG	ATTTGTTT	CAAGGGACTCG	ACTTGG	ATCGG	ACCTGCACCGG	1627	
Qy	1484	AGT	TGGCTCAG	TTTTTAC	GTCTCG	AGGACCG	AGTTTATG	CCATCT	TGATACTAA	1543	
Db	1628	TTGT	TGCATCTTCT	CTGCA	CGTGGTCT	TAATCC	CAGAGACG	CGGAGATTT	CTTAAACCGG	1697	
Qy	1544	ATAT	GATTTGCT	CCGGAGTCA	ATATCA	TTCATTTGG	GGCTTGGCC	TCAAAATCT	TAGGACCAACCG	1603	
Db	1688	ATG	TAATCG	CACCTGGTTGA	ATATTCTAG	CTGCTTGG	CCCTGAC	CGGATTTGGT	CTCCATCTG	1747	
Qy	1604	GACT	TCCTTATG	ATCA	AGAGAGTAA	ACTTTCACTG	TAATGT	CAGGAAC	TTCAATGCTTT	1663	
Db	1748	GTGT	ACTCTCTG	ATAAC	CGGAGAACTG	AGTTCA	CAATTTTAT	CAGGCAC	TTTCGATGGCGT	1807	
Qy	1664	GTCC	ACATGTTAG	CGGAATCA	CTGCTCTTAT	CCGGTCTG	CATACCC	GAACTGGT	CTCCAG	1723	
Db	1808	GCCCT	CAATGTTCT	GGTCTAG	CTGCTTGTCTT	AAAGCGGCT	CTATCC	CGAATTGG	AGCCCGG	1867	
Qy	1724	CTG	CAATCA	AATCCG	ATTGATG	ACAACAG	CGGATTTGT	TACGATCGT	CAAGGGAACGGA	1783	
Db	1868	CGG	GAATAAG	ATCAG	CAATGATG	ACCA	CGGCTTAC	ACGGTTG	TAAACAG	CGGTCGAGCCGA	1927
Qy	1784	TAA	AGGATGG-----	-----	TAA	CAAA	CCACCGCGGTG	TTTGGATG	TTGGAGCAGGGCATGGA	1837	
Db	1928	TGAT	TGGATG	AGTCC	CACTGG	CAATAC	GTCTTC	GGTAC	CGGAATATG	TTGTTCCGGT	1987
Qy	1838	ATCC	GCAAAAGG	CGGATATA	ACCGGGAT	TGGTTTAA	CAACATTT	CAAC	CACTGGAT	TACATAA	1897
Db	1988	ACCA	CCAGAGCTAT	GATCCG	GGACTAG	TTAC	GATATA	TAA	CATCTTAT	GATTAATCA	2047
Qy	1898	CTT	ACCTCTG	CAC	CTCTTTGG	ATTCA	AGATCAG	ATATTTT	TAGCAATCA	CTCATGAAG	1957
Db	2048	ACT	TTCTGTG	TAAATCT	CTAA	CTACTG	GAACCA	CAATTTG	TACGATAA	CCCGTCGCAAG	2107
Qy	1958	TGAG	CTGCA	TGGAA	TATTTCC	GGAATAA	ACCGGGTTT	TAGTCT	CAATTA	CTCCGTCGATAG	2017
Db	2108	CGGA	CTCG	ACGCTG	CGGAGG	CGAGCGG	GACAGCTG	CGGAAATTT	TGA	ACTTACCCGCTCGTTCT	2167

QY 2018 CCGTGATTTTCAA-----ACGTGGCAAGACTACGAGATGATCACAAGCGGTGTCA 2068
Db 2168 CTGTCGTGTTTCAGCAGTATGAGAGAGTAAATGTCGACACATTTTCATTAGGACTGTGA 2227
QY 2069 CTAAGCTGGGAGTCTTAACCTGATATCTACTAGTGAATGTCAAGCTCCAGAGGGGATCA 2128
Db 2228 CAAATGTAGCGGATTCGGATTCGGTTTACGAGATTAAAGATTAGGCGGCTAGAGGGACTA 2287
QY 2129 AAGTTATTGTCATCTTAAGAGACTTGTGTTCAAACAGTGTGATCAGACGCTGAGCTATA 2188
Db 2288 CAGTGACGGTTGAGCGGAGAGCTATCGTTTCAGACGGGTGGGCGAGAACTGAGTTTGTG 2347
QY 2189 GAGTATGGTTT 2199
Db 2348 TTGTTAGGTTT 2358

RESULT 11

AB213381
ID AB213381 standard; DNA; 2295 BP.

XX AB213381;

XX 21-JAN-2003 (first entry)

XX Arabidopsis thaliana stress regulated gene SEQ ID NO 1186.

XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX Arabidopsis thaliana.

XX WO200216555-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26685.

XX 24-AUG-2000; 2000US-227866P.

XX 26-JAN-2001; 2001US-264647P.

XX 22-JUN-2001; 2001US-300111P.

XX (SCRI) SCRIPPS RES INST.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Harper JF, Kreps J, Wang X, Zhu T;

XX WPI; 2002-304127/34.

XX Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses

XX Claim 144; SEQ ID NO 1186; 577pp + Sequence Listing; English.

XX The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
CC in methods of the invention.

CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.

XX Sequence 2295 BP; 553 A; 550 C; 586 G; 606 T; 0 other;

XX Query Match 17.1%; Score 398.2; DB 24; Length 2295;

XX Best Local Similarity 53.2%; Pred. No. 3,7e-113;

Matches 1072; Conservative 0; Mismatches 898; Indels 45; Gaps 9;
QY 205 TCTCGACTTCTCTACTCTATGGCTCTGGATTGAAGAGATTTCTGCTCAGTTGACTGAA 264
Db 166 TCTCGAATCGTCCATGTTTACCACACAGTCTTCCATGTTTCTCGGCGGTTTACTTCCA 225
QY 265 TCAGAAAGCGAGATATCGAGATATTCACCTGAAGTTTGTGACGTGAGACCTCACCATGTT 324
Db 226 GATGAAGCAGATAATCTCGGTAACCAACCCAGCAGTCTTGTGTTTTCGAAGACCGACGT 285
QY 325 CTTTCAGGTTCAACCACTTACTCTTACAGTTCTTGGGACTCGACGGTTTTCGAACACTCC 384
Db 286 CGAGAGCTTCACACCACAGTCTCTCTCAATTTCTTGGTTTACAA-----AACCAAAA 339
QY 385 GGTGTATGGTCTAAATCTCGGTTTGGTCAAGGCACAAATATCGGGCTGCTGATCTGGA 444
Db 340 GGAATAAGTCAAGATCTGATACGATCAGACGTAATCATTTGGCGTTTTCGACACCGGA 399
QY 445 GTTTGGCCTGAAAGTCTAGCTTTGACGATACCGGAATGCTTTCGATTCACCGGAATGG 504
Db 400 ATTTGGCGGAGCGGAGGAGTTTCTCAGATCTTAACCTCGGTCCTCAATTCCTCAAAAAGTGG 459
QY 505 AAAGGGAATTTCCAAAGAGAGAAATTTTCAGTTCTTCGAGGTGTAACCGGAGCTAATC 564
Db 460 AGAGGCGTTTTCGAATCCGAGGCCAGATTCAGTCTCGGAACCTGTAAACCTAAATATC 519
QY 565 GGTGTAGATTCTTCATCAGAGGACACCGTGTGCTAATTCACAGAGGAATCACCACAA 624
Db 520 GGAGCAAGATCTTCGTAAGGACAAACAGCGCT--GTAATCGGAGGAATCAACAAA- 576
QY 625 ATGCTCGTGAATACATTTCCGCAAGAGATTTCAACGGGACACGGGACTCACACCGCTCA 684
Db 577 ---ACCGTTGAGTTTCTATCTCTCTGAGCGCGATGGACACGGTACTCACACTTCTCA 633
QY 685 ACAGTTGGTGGATCCTCTGTTTCGATGGCGAATGTTCTTTGGCAATGGAGCTGTGTGGCT 744
Db 634 ACCGCCCTGGCGCTCACGCTTTTAAAGCGAGTATGTCGGTTTACGCTCCGCTGTAGCC 693
QY 745 CFTGGGATGCTCTCTGAGCTCACATTCGAGTCTATATAAGTCTGTGTTTCAAT---GGT 801
Db 694 AAAGGTGTTGCTCCAAAGCTCGTATCGCGCGCTTACAAAGTCTGTGGAAGATTCGGGT 753
QY 802 TGTTACAGCTCTGACATTTAGCAGCTATAGATGTAGCGATTTCAAGATAAGTCGATGTT 861
Db 754 TGTCGATTCGATATTTCTCGCGCTTTGATGCGCGTGTAGAGACGCTGTGCGAGTT 813
QY 862 CTTTCGCTTTCCCTTGGCGTT-----TCCCTATTCTCTTGTATGATGACACAATC 912
Db 814 ATATCGATCTCAATCGGTGCTGGAGACGGGATTAATTCGCGCTATTACCTCGATCCAATC 873
QY 913 GCCATTGGACATTCGAGCCATGGACGGGTATATCTGTAATCTGTGAGCTGTTAAC 972
Db 874 GCTATAGGCTCGTACGGCGCGCTCGAAAGGAATCTTCGCTCTCTCTGCGCGGAAC 933
QY 973 AACGTCCTCAATCGAAAGCTCTGTTGCAACACACAGCTCTTTGGGTCTCAACATTGGCGCA 1032
Db 934 GAAGAGCTTAACGATGTGTCAGTTACGAACTCGCGCGTGGGTAAACACCGTTGGTGTCT 993
QY 1033 GGCACGCTTCATCGAAGATTTCCGCTGTGTCAGATTAGCCAAACGGAAGAGTCTCTAT 1092
Db 994 AGTCAATCGATCGAATTTCCAGCGGATGCTATTCTCGCGACGAGACATCGTCTCAGA 1053
QY 1093 GGAGAGCTATGTTATCCGGGAAAGGATTAAGATCCCGGAGAGAGGTGAGGTGATT 1152
Db 1054 GGAGTGTCTCTTTACGCTGGAGTACCTTTAAACCGTCTGTATGTTTCCGGT---GGTTTAT 1110
QY 1153 TACGTCACAGAGAGATTAAGGAGTGTCTCTGTTTGGAGGGTCACTTCCAGAGAA 1212
Db 1111 CCCGTAATCGGGAATGTCTATGATGAGGACACGCTGTATGCCAAG 1170
QY 1213 GAAATCCGAGGCAAAATGCTGATTTGTGATCGCGAGTCAATGGAAGATCGGAAGAGGA 1272
Db 1171 CAAGTGGGGTAAATAGTAACTCTGATAGAGAGACGCTCCACGCGTAGCCAAAGGA 1230

Qy	1273	GAACGGTTAAAGAAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCAA	1332
Db	1231	TTGGTTGTGAAGAAAGCAGGTGGTGTCCGAATGATCTCGCTAATGAGCATCTTAACGGT	1290
Qy	1333	GAAGAAGATCTATTGAGGTTCACTCTTACCAGCTACATGATGGTTGTTACACTGATCA	1392
Db	1291	GAAGATTATGCGGAGATGCTCATCTTATTCACGCTGTCGGTTGGATCAAAAGGAAGGA	1350
Qy	1393	GTCTTTCTGAAGGCTTATGTAAATGCCACGCTGAACCAAGGCGCGATAATTTTGGT	1452
Db	1351	GATAGAAATCAAAGCATATGCTTCTTCACATCCGAATCCAAATGGTTCAATTGATTCAGA	1410
Qy	1453	GGTACGCTGATTTGGAGGTCACGAGCACCGAGGTGGCTCAGTTTTCAGTCTCGAGGACCG	1512
Db	1411	GGAACATAGTTGGGATTTAAACCGCTCCGGTTATTGCTTCTTCGGTAGAGGACCA	1470
Qy	1513	AGTTTAGCCAAATCCTTCGATACTAAACCCGAGATGATTCGTCGGGAGTCAATATCAT	1572
Db	1471	AACGGTTAAAGCCCGAGATTTCTTTAAACCGGATTTGATTCGCCGAGTTAAATCCTC	1530
Qy	1573	CGCGCTTGGCTCAAAATCTAGGACCAACCGGACTTCCTTATGATTTCAAGAAGATTAA	1632
Db	1531	GCCGCATGGACAGACGCTGTTGGACCTTACAGTTTGGCTCAGATCCAAAGAAACCGAA	1590
Qy	1633	TTCACTGTAAATGTCAGGAACTTCAATGCTTTGTCCACATGTTAGCGGAATCACTGCTCT	1692
Db	1591	TTCAACATTTCTCCCGTACTTTCATGGCATGCTCTCAGTTAGTGTGCGGCGCGCTT	1650
Qy	1693	ATCCGCTGTGCATACCGCAACTGTGCTCCAGTCCAAATCAAAATCCGCAATGATGACAACA	1752
Db	1651	CTCAATCCGCTCATCCAGATTGGAGCCCTGCCGTGATACGATCGGCAATGATGACAACG	1710
Qy	1753	CGCGATTGTACGATCGTCAAGGAAAGCGATAAGGATGGTAAC-----AAACAGCC	1806
Db	1711	ACTAACCTCGTCGATAACTCTPAAACCGCTCGTTGATCGATGAATCCACCGGGAATCGGCT	1770
Qy	1807	GGTGTGTTTGGATTGGAGCAGGCGATGTGAATCCGCAAAAGGCGATAAACCCGGGATG	1866
Db	1771	ACGCTTATGATTAACGGGTGGGTGATTTAAATTTGGCGCGGTATGAATCCGGGTCTT	1830
Qy	1867	GTTTACAACATTC AACCAGTGGATTACATAACTTACCTCTGCATCTTGGATTCAACAAG	1926
Db	1831	GTCTACGATATAACTAATGATGATTAATACGTCTTCTTTGCTCCATCGGGTACGACCA	1890
Qy	1927	TCAGATATTTAGCAATCACTCATAGAAAGTGAGTGCATGCAATGGAATATTGCGGAAAAAC	1986
Db	1891	AAGACGATCCAAAGTGATAAAGAAACACCGGTGAGATG--TCGACGACCAAGAAACCG	1947
Qy	1987	CCGGTTTTAGTCTCAATTAACCGTCATAGCCGTGATGATTTTC-----AAACGTGGC	2037
Db	1948	TCTCCCGGGAATTTGAATATCTCTTCGATCAGCGCGGTGTTTCTACTAATAAGAGGGA	2007
Qy	2038	AAGACTACGGAGATGATCACAAGCGGTGTCACTAACGTTGGAGTCTCTAATCGATATAC	2097
Db	2008	TTGTTGATGAATAAATCTTTTATAGGACCGGACGAATGTCCGGCAGGCTAGGCGGTTTAT	2067
Qy	2098	TCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCAATCTTAAGAGACTTGTG	2157
Db	2068	CGGCGGAGGATAGAGTCGCGCAGGAGGAGTGACGCTGACAGTGAAACACACTGACCTTG	2127
Qy	2158	TTCAACAACGTTGGATTCAGACGCTGAGCTATAGAGT	2192
Db	2128	TTTACTTTCCGCGGTTTAAGAGACGGAGCTATGCGGT	2162

RESULT 12

RESUL 12
AAC42230

AAC42230
ID AAC42230 standard; DNA: 2489 BP.

1000

AAC42230;

XX

DT 17-OCT-2000 (first entry)

XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 34769.
DE	Hybridisation assay; genetic mapping; gene expression control;
XX	protein identification; signal transduction pathway;
KW	metabolic pathway; promoter; termination sequence; ss.
KW	
XX	
OS	Arabidopsis thaliana.
XX	
XX	EPI033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-01211825.
PR	05-MAR-1999; 99US-01231190.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128214.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.

2

QY 505 AAGAGGATTTGCAAGAGGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAGCTAATC 564
Db |||||
654 AGAGGGCTTTGCGAATCCGAGGACGATTCAGTTCCTCGGAACTGTAAACCGTAAATATC 713
QY |||||
565 GGTGCTAGATTTCTCATCAGAGACACCGTGTGCTTAATTCACACAGAGAAATCACAAC 624
Db |||||
714 GGAGCAAGATTTCTCGCTAAGGACAAACAGCGCT- -GTAATCGAGGAATCAACAA- 770
QY |||||
625 ATCCCTCGTGAATACATATTCGCAAGAGATTCACGGGACACGGGACTCACACCGCCTCA 684
Db |||||
771 - - -ACGTTGAGTTTCTAICTCTCTGTCGCGCATGGACACGGTACTCACACTTCCTCA 827
QY |||||
685 ACAGTTGGTGGATTCCTGTTTCGATGGCAATGTTCTTGCAATGGAGCTGGTGGCT 744
Db |||||
828 ACCGCGCTGGCGCTCAGCGCTTTTAAAGCGAGTAIGTCCGTTACGCTTCGCGTGAAC 887
QY |||||
745 CGTGGAGTGGCTCTCGAGCTCATATGCAIGCTATAAAGTCTGTGGTTCAT- -GGT 801
Db |||||
888 AAAGGTGTTGCTCCAAAGCTCGTATCGCGCCTACAAAGTCTGTTGGAAGATTCGCGT 947
QY |||||
802 TGTTACAGCTCTGACATTTAGCAGCTATAGATGTAGCGATTCAAGATAAGTTCGATGTT 861
Db |||||
948 TGTTGATTCGATATTTCTGCGCGCTTTGATGCGCTGTAGACAGCGTGTGACGCT 1007
QY |||||
862 CTTTCCGCTTTCCCTTGGCGGTT- - - - -TCCCTATTCTTTGTTATGATGACACAATC 912
Db |||||
1008 ATATCGATCTCAATCGTGTGGAGACGGGATTACTTCGCGTATTACCTCGATCCAATC 1067
QY |||||
913 GCCATTGGAAATTCGAGCCATGGAACCGGCTATATCTGTAATCTGTCAGCTGTAAC 972
Db |||||
1068 GCTATAGGCTCGTACGGCGCGCTCGAAAGGAATCTTCTGTTCTTCTCTCGCGGAAAC 1127
QY |||||
973 AACGGTCCAATCGAAAGCTCTGTTGCAACACAGCTCTTGGTCTCAACCAATGGCGCA 1032
Db |||||
1128 GAAGGACCTAACCGGTATGTCAGTTACGAACCTCGCGCGTGGGTAAACCAACGTTGGTGT 1187
QY |||||
1033 GGCACGCTGATCGAAGATTTCCGCTGTGTCAGATTAGCAACCGGAAGCTTCTCTAT 1092
Db |||||
1188 AGTACAATCGATCGGAATTTCCAGCGGATGCTATTCTCGGCGACGACATCGTCTCAGA 1247
QY |||||
1093 GGAGAGTCATTGTTATCCGGGAAAGGTATAAAGATTCGCGGAGAGAGAGTTCAGGTGATT 1152
Db |||||
1248 GGAAGTCTCTTACGCTGGAGTACCTTTAAACGCTCGTATGTTTCCGCT- - -GGTTTAT 1304
QY |||||
1153 TAGCTCACAGGAGGATAAAGAAAGTGTCTGTTTGGAGGGTCACTTCCAGAGAA 1212
Db |||||
1305 CCGGTAATTCGGGAATGTATCAGCGTCTCTATGTATGGAGAACACACGCTTGATCCGAAG 1364
QY |||||
1213 GAATCCGAGGCAAAATGGTATTTGTATCGCGGAGTCAATGGAAGATCGGAGAAAGGA 1272
Db |||||
1365 CAAGTGGGGTAAATAGTAATCTCGATAGAGGAAGCAGTCCACGCTAGCCAAAGGA 1424
QY |||||
1273 GAAGCGTTTAAAGAGCTGAGGAGTTGCAATCATCTTAGCCCAATACAGAGATCAACCA 1332
Db |||||
1425 TTGGTTGGAAGAACAGAGTGTGTCGGAATGATTTCTGCTAATGGAGCATCAACGCT 1484
QY |||||
1333 GAAGAGATTTCTATTGACGTTCTATCTTTACCAGCTTACATTTGATTTGGTTACACTGATCA 1392
Db |||||
1485 GAAGGATTAGTCGAGATGCTCATCTTATTCAGCCTGTGCGCTTGGATCAACAGAGGA 1544
QY |||||
1393 GTCCTCTGAAGGCTTANGTTAATGCCAGCGTGAACCAAGCGCGGAGTAATTTTGGT 1452
Db |||||
1545 GATAGAATCAAGCATATGTTCTTTCATATCCGAATCCAAATGCTTCAATTGATTTTCA 1604
QY |||||
1453 GGTACGTTGATTTGGAGTTCAGGACCGCGAGTGGCTCAGTTTTCAGCTCGAGGACG 1512
Db |||||
1605 GGAACATAGTTGGGATTAACCGGCTCCGGTTATTGCTTCTTCTCCGGTAGAGAACCA 1664
QY |||||
1513 AGTTTACCAATCTTCTGATCTATAAACCGGATATGATTTGCTCCGGAGTCAATATCAT 1572
Db |||||
1665 AACGGTTAAGCCCGGAGATCTTAAACCGGATTTGATTTGCTCCCGGAGTTACATCTCTC 1724

QY 1573 GCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAGAAGAGTTAAC 1632
Db |||||
1725 GCGCATGGACAGACGCTGTTGGACCTTACAGGTTGCGGTCAGATCAAGGAAACCGAA 1784
QY |||||
1633 TTCACTGTAAATGTCCAGAACTTCAATGCTTTGTCCACATGTTAGCGGAATCACTGCTCTT 1692
Db |||||
1785 TTCAAAATCTTCCCGTACTTCAATGGCAATGCTTCACTGATTTAGTGGCGGCGCTT 1844
QY |||||
1693 ATCCGCTCTCATACCCGAACTGGTCTTCCAGCTGCAATCAAAATCCGCAATGATGACAACA 1752
Db |||||
1845 CTCAAAATCCGCTCATCCAGATTTGGAGCCCTGCCGTGATCGATCGGCAATGATGACAACG 1904
QY |||||
1753 GCGGATTTGATCGATTCGTAAGGAAAGCGATAAAGATGTTAAC- - - - -AAACGAGCC 1806
Db |||||
1905 ACTAACTCTGTCGATACTCTAAACCGCTCTTATCGATGATGATCCACCGGAAATCGGCT 1964
QY |||||
1807 GGTGTTTTCGATTTGAGGAGGCGCATGCAATCCGCAAGGCGATAAACCGGAGTTG 1866
Db |||||
1965 ACGCTTATGATTTACGGGTCGGGTCATTTAAATTTGGGCGGGCTATGAATCCGGGCTT 2024
QY |||||
1867 GTTTACAACATTCACCCAGTGGATTACATAACTTTACCTCTGCACCTCTTGGATTCAAGA 1926
Db |||||
2025 GTCTACGATATACTAATGATGATTACATTACGTTCTTGTCTCCATCGGTCAGGACCA 2084
QY |||||
1927 TCAGATATTTAGCAATCACTCTCAAGAACGTCGAGTGCATGGAATTTGCGGAAAC 1986
Db |||||
2085 AAGACGATCCAAAGTGAFAACAAGAACACCGGTGAGATG- - -TCCGACGACCAAGAAACCG 2141
QY |||||
1987 CCGGTTTGTAGTCTCAATTAACCGTTCGATAGCGGTGATTTTC- - - - -AAACGTGGC 2037
Db |||||
2142 TCTCCCGGAAATTTGAATTTATCTTCGATCACCGCGGTGTTTCTACTAATAGAAGGA 2201
QY |||||
2038 AAGACTACGAGATGATCAACAAGCGGTGTCATAACGTTGGAGTCTTAACCTCGATATAC 2097
Db |||||
2202 TTGTGATGTAACCTGTTTATAAGAACGCGACGAATGTCCGGCAGGCTGAGCGGTTTAT 2261
QY |||||
2098 TCACTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCATCTTAAGAGACTTGTG 2157
Db |||||
2262 CCGGCGGAGTAGAGTCCCGAGGAGTACGCTGACAGTGAACCAACCTAGGCTTGTG 2321
QY |||||
2158 TTCAAAACGCTGGATCAGACGCTGAGCTATAGAGT 2192
Db |||||
2322 TTTACTTCCGCGGTTAAGAGACGAGCTATGCGGT 2356

RESULT 13

AAC32938
ID AAC32938 standard; DNA; 1653 BP.

AC AAC32938;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 1197.

XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128734.
PR 16-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 26-APR-1999; 99US-0131449.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152163.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0158369.

```
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160788.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.2%; Score 285; DB 21; Length 1653;
Best Local Similarity 53.1%; Pred. No. 7.7e-78;
Matches 684; Conservative 0; Mismatches 595; Indels 18; Gaps 3;

QY 962 CAGTGGTAAACAGCGTCCAAATCGAAGCTCTGTGTGCAACACAGCTCCTTGGTCTCAA 1021
DB 2 CTGCTGTAATAGTGGTCTCTACTAGAGCTTCTGTGTGTAATGTTCCTTGGTGTATGA 61

QY 1022 CCATTGGCGGACCGCTTCATCGAAGATTTCCGCTGTGGTGCAGATTAGCCAAACGGAA 1081
DB 62 CTGTTGGTGGTACTTTAGATAGAGATTTCCGGCTTTTGGCAATCTCGGTAAACGGGA 121

QY 1082 AGCTTCTCTATGGAGAGTCATTGTATCCGGAAAGGTATAAAGATGCCGGGAGAGAGG 1141
DB 122 AACGACTTACCGGTGTTTCGTGTATAGCGGTGTGGGAATGGGA-----CGAAGCCGC 175

QY 1142 TTGAGGTGATTACGTACAGAGAGAGATAAGGAAGTGTCTGTGTTGAGAGGTCAC 1201
DB 176 TTGAATTGGTTTATAATAAAGGAATAGTAGTTCGAGTAATCTTTGTTTACCTGGTTCGC 235

QY 1202 TTCAGAGAGAAATCCGAGGCAAAATGGTGATTTGTGATCGCGGAGTCAATGGAAGAT 1261
DB 236 TTGATTCGAGTATTGTTCCGTGGGAAGATTGTTGTTGTGATAGAGGTGTTAAATGCTAGG 295

QY 1262 CGGAGAAAGGAGAGCGGTAAAGAGCTGAGAGAGTTGCAATGATCTTAGCCCAATACAG 1321
DB 296 TTGAGAAAGGAGCTGTGTTTACAGATCTGTGTGTTTGGGATGATTAATGGCGAATACG 355

QY 1322 AGATCAACCAAGAGAGAGATTCTATTGACGTTTCATCTCTTACCAGCTACATTTGGTT 1381
DB 356 CTGCGAGTGGAGAGAGCTTGTGGCGGATGATCAITTCCTCCCGCATGCTGTAGGGA 415

QY 1382 ACACGTAGTCACTCTCTGAGAGGTTATGTTAATGCAACGCTGAAACCAAGCGCGGGA 1441
DB 416 AGAAGACTGGTGATTACTTAGGGAGTATGTTAAGTCAGATTCTAATCCAACTGCTCTTC 475

QY 1442 TAAATTTTGGTGTACGTTGATTTGGAGGTACAGAGCACCGGAGTGGCTCAGTTTTCAG 1501
DB 476 TTGTTTTTAAAGGAACGGTCTTTGACGTTAAGCCGCTCTCCTGTGTTGCTGCTTTAGCT 535

QY 1502 CTCGAGGACCGAGTTTAGCCATCTCTCGATACCTAAAAACCGGATGATTTCTCCCGGAG 1561
DB 536 CGAGAGGTCCTAATACTGTTACTCTCTGAACTTGAAGCCTGATGTTATTTGGTCTCGAG 595
```

```
QY 1562 TCAATATCATTTGGCGCTTGGCCTCAAATCTAGGACCAACGGGACTTCTTATGATCAA 1621
DB 596 TGAACATTTTGGCTGGTGGTCTGACGCTATTGGTCTCTACTGGTCTTGACAGGACTCTA 655
QY 1622 GAAGAGTTAACTTCACTGTAATGTGAGAACTTCAATGTCTTGTCCACATGTTAGCGGAA 1681
DB 656 GGAGACTCAGTTTAAACATCATGTGAGTACGTCAATGTCAATGCCACACATCAGTGGTT 715
QY 1682 TCATGCTCTTATCCGGTCTGCAACCGGAACCTGGTCTCCAGCTGCAATCAATCCGCAAT 1741
DB 716 TAGCGGTCTTTTGAAGACAGCTCACCTGAGTGGAGTCCGAGTGTCTATCAATCAGCTC 775
QY 1742 TGATGACAAACAGCGGATTTGTACGATCTCAAGGGAAGCGATAAAGATCGT-----A 1795
DB 776 TCATGACTACAGCTTACGTTCTTGACAAACCAACGCTCTCTCCATGATCTCGACACA 835
QY 1796 ACAACACAGCGGTGTGTGTTGCGATTGGAGCAGGCGCATGGAATCCGCAAAAGCGCATAA 1855
DB 836 ACAGCCTATCTTAACCCATATGCTCATGGCTCGGCGCATGTAGATCCCAAAAGGCTCTCT 895
QY 1856 ACCCGGGATTGGTTTACCAACATTTCAACAGTGGATTACATAACTTACCTCTGCACTCTTG 1915
DB 896 CACCAGGTCTTGTCTACGACATCTCAACCGAGGAATACATCAGGTTTTTGTGCTCTCTAG 955
QY 1916 GATTCAACAGATCAGATATTTTAGCAATCACTCATPAGAACGTGAGCTGCAATGGAATAT 1975
DB 956 ACTACACAGTCGATCACATTTGTTGGAT-----TGTGAAGCGACCTAGCGTTAACTGCT 1009
QY 1976 TCGCGAAAAACCCGGGTTTTAGTCTCAATTTACCGTGCATAGCCGTGATTTTCAAACGTG 2035
DB 1010 CGAAGAAAGTCTCAGATCTGTCAGCTCAACTACCCAAAGTTTTCGGTTTTTGTGGGG 1069
QY 2036 GCAAGACTACGGAGATGATCAAGGCGGTGTCACCTAACGTTGGGAGTCTCTAACTCGATAT 2095
DB 1070 GTAAGAAGTGTGCGGTACACTCGGAAGTAAACAAATGTTGGTGCAACAGCTCGGTTT 1129
QY 2096 ACTCAGTGAATGCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCAATCTCTAAGAGACTTG 2155
DB 1130 ACAAGTGAAGCTTAAATGGAGCTCTAGTGTGGAATCTCTGTAAACCATCGAAACTTT 1189
QY 2156 TGTTCAAACACGTGGATCAGCGCTGAGCTATAGAGTATGTTTGTATTTGAAGAGAAA 2215
DB 1190 CGTTTAAAGCGTGGGAGAGAGAGGTACACAGTCAAGTTTGTAGCAAGAAAGAG 1249
QY 2215 ACAGAGGAGGAGGTGGCTAGCTTTG 2242
DB 1250 TGAGTATGACGAACAAAGGCTGAGTTG 1276

RESULT 14
AAC42099
ID AAC42099 standard; DNA; 2402 BP.
XX
XX AAC42099;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 34284.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX BP1033405-A2.
XX
XX 05-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
PR
```

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135829.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 13-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154773.
PR 22-SEP-1999; 99US-0155133.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

```
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          9.8%; Score 228.2; DB 21; Length 2402;
Best Local Similarity 48.3%; Pred. No. 5.6e-60;
Matches 977; Conservative 0; Mismatches 963; Indels 81; Gaps 9;

QY 211 CTTCTCTACTCTTGTGCTTCGAGTTGAAGGATTTCTCTCAGTTGACTGATCAGAA 270
DB 237 CTTCTCCACACATATAAGCATGGCTTCTCAGGTTTTCAGCTCGTTTGACAGAGAG 296

QY 271 GCCAGATACAGATATTCACCTGAAGTTGTTGCACTGAGACTGACCATGTTCTTCAG 330
DB 297 GCCAAGTCATAGCAGAAACCGGAGTGTGTTTCAGTTTTCAGCCACACTTCCAG 356

QY 331 GTTCAAACCACTTACTCTTCAAGTTCTTGGGACTCGACGGTTTGGAAACTCCGGTGA 390
DB 357 CTTCTACAACCTCATTCAGGACTTTCTCAAGTACCAACATCAGTAAAGGTCGATTC 416

QY 391 TGGTCTAAATCGGTTTGGTCAAGGCACA-----ATTATCGCGTGTGTGATACT 441
DB 417 GGTCCACCTTCATCAGCCTCAGATGGAKCATATGATAGCATTTGCGGAATTTCTGACA 476

QY 442 GGAGTTGGCTGAAAGTCTTAGCTTTTGACGATACCGGAATGCTTCGATTCACCGGAA 501
DB 477 GGGATATGGCCAGAGTCAGAGAGTTTCAATGACAAGACATGGGTCCAATTCGCTCGG 536

QY 502 TGGAAAGGGATTTGCCAAGAGGAGAAAGTTTCAGTTTCTTCGAGCTGTAAACCGGAAGCTA 561
DB 537 TGGAAAGGTATCATGATGGAGCAAGGACTTCAAGTCTTCCAACTGTACAGAAAGATC 596

QY 562 ATCGGTGCTAGATTCTTCATCAGAGGACACGGTGTGCTAATTCACAGAGGAATACCA 621
DB 597 ATTGGAGCAAGATACTACA-----AAAATCCT 623

QY 622 AACATGCTCTGTGATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACCCGCC 681
DB 624 GATGATGATTCAGAAATACTATACCACAAGGGATCTCATCGGTACAGGTTCTCATGTGTC 683

QY 682 TCAACAGTTGTTGATCCTGTTTCGATGCGGAATGTTCTTGGCAATGAGGTGGTGTG 741
DB 684 TCCACCATAGCTGGATTCGCCGTGGAGAAATGCTTCCTACTACTATGTTAGCTTCGGGACT 743
```

```
QY 742 GCTCGTGGATGGCTCTCTGGAGCTCATTGCGAGTCTATAAAGTCTGTTGGTTCATGGT 801
DB 744 GCAAAGGGAGGTTTACAAAACGCTAGAAATCGCTATGTACAAAGTATCAATCCAGGGGA 803

QY 802 TGTACAGCTCTGACATTCAGCAGCTATAGATAGCTAGCATTCAGATTAAGATAAGTCGATGTT 861
DB 804 TGCACGTGGCTCTCTATCTATCTTAGCTGCTTTTGATGATGCAATCGCAGATGGAGTTGATGTT 863

QY 862 CTTTCGCTTTCCCTTGGCGGTTTCCCTATTCT-----TTGTATGATGACACAATC 912
DB 864 CTAICTCTGCTCTTTGGAGCTCCAGCATACGCTCGCATCGACTTGAACACTGATCTATT 923

QY 913 GCCATTGGAACATTCGAGCCATGGAACGGGTATATCTGTAATCTGTGACGTGCTGAGCTGGAAC 972
DB 924 GCCATTGGAGCGTTTCAAGCGGTGGAGCAAGGAATCTTGGTGTATCTGCTCTCGCGGTAAT 983

QY 973 AACGGTCCAATCGAAAGCTCTGTTGCAACACACTCTCTGGGTCTCAACCATTTGGCGCA 1032
DB 984 GATGGACCTGATGCGGTTACAGTTACTAATACTGCACCTTGGATAATGACCGTTGCTGCC 1043

QY 1033 GGCACGCTTGATCGAAGATTTCCGCTGTGTGATAGATTAGCCAAACGGAAAGCTTCTCTAT 1092
DB 1044 AACACTATTGATAGACTTTTGTGATCTGTGATCTAGCGGCAATAAGTCAATCAAG 1103

QY 1093 GGAGAGTCATTGTATCCGGGAAAAGGTATAAA-----GAATGCCGGGAGAGAGGTT 1143
DB 1104 GGTGAAGGTAACACTTTTCAACGTTAGTAAATCTCTGTGTATCTCTGATTCAATGCGC 1163

QY 1144 GAGGTGATTTAGTCAAGGAGGAGATAAGGAAGTG---AGTTCGTTTGGAGGGTCA 1200
DB 1164 AAGTCTGCTAAGAGCGCTGATGATCATGAAAGGATTCAGCCAGGCGCTGTGACTGTGATTTCT 1223

QY 1201 CTTCAAGAGAGAAATCCGAGGCAAAATGTTGTTGATCGCGAGTCAATGGGAAGA 1260
DB 1224 CTAGATCAAGAGAGAGTTAAAGGAAGATTTGTTATGCGAGAAAGTTGGTGGATCATAT 1283

QY 1261 TCGGAGAAAGAGAGAGCGGTTAAAGAAAGCTGGAGAGTTGCAATGATCTTAGCCAAATACA 1320
DB 1284 TATGATCATCTCGCTAGGACAAAGTGAAGAGCAAGAGGATCTAGTTGCTTTGTA 1343

QY 1321 GAGATCAACCAAGAAAGATTTCTATTGACCTTCACTCTTTACAGCTACATTTGATGTT 1380
DB 1344 GATGACAGAACTAGAGCAGTTGCTAGTCTTATGTTAGTCTTCTACTACCGTAATTGAC 1403

QY 1381 TACACTGATCAGTCTTCTGAGGCTTATGTTAATGCCACGGTGAACCAACGAAGCGCGG 1440
DB 1404 TCAAAGGAAGCAGCTGAGATCTTCTTACCTCAACTCAACAAAGATCTCTGTTGCAACA 1463

QY 1441 ATAAATTTTGGTGGTACGGTGAATGGGAGGTCAAGAGCACCGGAGGTGGCTCAGTTTCA 1500
DB 1464 ATTCTTCCACTGCAACAGTTGAAAGTTCACTGCTGCTGCTGTTGCAATATTTTCT 1523

QY 1501 GCTCAGGACCGAGTTTAGCCAATCCTTCGATACATAAACCAGGATATGATTTGCTCCGGGA 1560
DB 1524 TCCAGAGACCTTCAAGCCTCAAGAAGCATTTCAAACCTGACATTTACCGCACCGAGA 1583

QY 1561 GTCAATATCATTTGGGCTTGGCTCAAAATCTAGACCAACCGGACTTCTTATGATTTCA 1620
DB 1584 GTCTCGATACTCGCTGCTGATGACTGGAA---CGACTCAAGCATTTCTCTGGAAGGCAAG 1640

QY 1621 AGAAGAGTTAACTTCACTGTAATGTCAGAACTTCAATGTTCTGTCACATGTTAGCGGA 1680
DB 1641 CCGGTTTCTCAGTATAACGTATATCAGAACTTCCATGGGAGCTCTCTATGTTTCAGCT 1700

QY 1681 ATCACTGCTCTTATCCGGTCTGCAATCCCGAACTGGTCTCCAGCTGCAATCAAAATCCOCA 1740
DB 1701 GTTGATCTCTGATCAAAATCAGCATCCACATGGGGTCCATCCGCGATCAGATCAGCA 1760

QY 1741 TTGATGACAAACAGCG---GATTTGACGATCGTCAAGGGAAGGATTAAGATGCTGAAC 1797
DB 1761 ATTATGACAAACAGCGACTCAAACAAACGACAAAGGCTTTATAACACACAGAACTGGT 1820

QY 1798 AAACACCGCGGTGTTTTCGATTTGGAGCAGGGGATGTAATCCGCAAAAGCGGATAAAC 1857
```


us-09-806-767-1.rng

Tue Feb 3 10:49:41 2004

Job time : 460 secs

GenCore version 5.1.6
Copy-right (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 21:38:00 ; Search time 119 Seconds
(without alignments)
8634.789 Million cell updates/sec

Title: US-09-806-767-1
Perfect score: 2328
Sequence: 1 atggaccacaaacctttctt.....taacctgaagactaactga 2328

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	62	2.7	7218	1	US-08-232-463-14
2	58	2.5	293	4	US-09-313-294A-6881
3	44.8	1.9	1236	4	US-09-445-472-2
4	44.8	1.9	1566	3	US-08-894-818B-4
5	44.8	1.9	1962	3	US-08-894-818B-34
6	44.8	1.9	1962	4	US-09-445-472-15
7	44.4	1.9	295	4	US-09-313-294A-7407
8	37.2	1.6	289	3	US-09-007-005-17
9	37.2	1.6	289	3	US-09-244-796-17
10	37.2	1.6	3489	4	US-08-728-323A-1
11	37.2	1.6	3489	4	US-09-298-568-1
12	37.2	1.6	3489	4	US-09-410-399-1
C 13	37.2	1.6	32207	2	US-08-770-379-20
C 14	37.2	1.6	32207	2	US-08-757-669A-20
C 15	37.2	1.6	32207	4	US-09-230-371A-20
C 16	37	1.6	1230025	4	US-09-198-452A-1
17	36	1.5	1859	3	US-08-894-818B-15
18	36	1.5	1977	3	US-08-894-818B-2
19	36	1.5	1977	3	US-08-894-818B-6
20	36	1.5	1977	4	US-08-445-472-11
21	36	1.5	2595	4	US-09-281-766-18
22	36	1.5	7218	1	US-08-232-463-14
C 23	34.6	1.5	938	4	US-09-205-258-211
24	34.4	1.5	1146	2	US-08-504-365B-74
25	34.4	1.5	1494	3	US-09-255-502-1
26	34.4	1.5	1496	6	RE34606-1
27	34.4	1.5	1497	1	US-08-322-677A-6

28	34.4	1.5	1497	1	US-08-322-676-6	Sequence 6, Appli
29	34.4	1.5	1497	3	US-08-898-218-6	Sequence 6, Appli
30	34.4	1.5	1497	3	US-08-848-793-6	Sequence 6, Appli
31	34.4	1.5	1497	4	US-09-445-270-1	Sequence 1, Appli
32	34.4	1.5	1497	4	US-09-178-173A-1	Sequence 1, Appli
33	34.4	1.5	1497	4	US-08-322-678-6	Sequence 6, Appli
34	34.4	1.5	1497	6	5472855-1	Patent No. 5472855
35	34.4	1.5	1513	4	US-09-178-155-1	Sequence 1, Appli
36	34.4	1.5	1868	1	US-08-069-863-1	Sequence 1, Appli
37	34.4	1.5	1868	1	US-08-309-069-1	Sequence 1, Appli
38	34.4	1.5	1868	1	US-08-562-833-1	Sequence 1, Appli
39	34.4	1.5	1868	4	US-09-659-749-1	Sequence 1, Appli
40	34.4	1.5	1868	4	US-09-672-105-1	Sequence 1, Appli
41	34.4	1.5	1868	5	PCT-US95-05520-1	Sequence 1, Appli
42	34.4	1.5	8119	1	US-08-460-343B-1	Sequence 1, Appli
43	34.4	1.5	8119	1	US-08-398-028B-1	Sequence 1, Appli
44	34.4	1.5	8119	2	US-08-504-265B-1	Sequence 1, Appli
45	34.2	1.5	1867	1	US-07-772-087-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14
Query Match 2.7%; Score 62; DB 1; Length 7218;

```
Best Local Similarity 4.9%; Pred. No. 3.8e-09;
Matches 20; Conservative 229; Mismatches 159; Indels 0; Gaps 0;

QY 1095 AGAGTCATGTATCGGGAAAGGTATAAGAAATCCCGGAGAGAGGTGTAGGTGATT 1154
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1449 AGAGAAATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1390

QY 1155 COTCACAGGAGAGATAAGGAAGTGAGTTCTGTTTGTAGAGGGTCACCTCCAAGAGA 1214
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1389 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1330

QY 1215 AATCCGAGGCAAAATGGTGAATTTGTATCGGAGTCAATGAAGATCGGAGAAGAGA 1274
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1329 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1270

QY 1275 AGCGGTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCAATACAGAGATCAACCAAGA 1334
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1269 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1210

QY 1335 AGAAGATTCATTGAGGTTCACTCTTACCAGCTACATTTGTTTACACTGAGTCAGT 1394
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1209 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1150

QY 1395 COTTCGAGGCTATGTTAAATGCCACGGTGAACCAAGCGCGGATAATTTTGGTGG 1454
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1149 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1090

QY 1455 TAGCGTGTATGGAGTCTACGAGCAGCGGAGTGGCTGCTTTCAGC 1502
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1089 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1042

RESULT 2
US-09-313-294A-6881
; Sequence 6881, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6881
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700352577H1
US-09-313-294A-6881

Query Match 2.5%; Score 58; DB 4; Length 293;
Best Local Similarity 54.3%; Pred. No. 7.6e-09;
Matches 140; Conservative 0; Mismatches 115; Indels 3; Gaps 1;

QY 622 AACATGCTCGTGATACATTTCCGAGAGATTCACGGGACACGGGACTCACCGCC 681
Db 5 AACGCATCTCAGGATCTTGCTTACCATCTGATGATGGCCATGGGACGCATACAGCA 64

QY 682 TCAACAGTTGGTGGATCCTCTGTTTCGATGGCAATGTTCTTGGCAATGGAGTGGTG 741
Db 65 TCCATTGCTGTGCAATCATGNAATCTCTGTTGTTGTGCTGCAATCAATTTGGAAT 124

QY 742 GCTCGTGGGATGGCTCTCTGGAGCTACATTCAGTCTATAAGTCTGTGTTGTTCAA--T 798
Db 125 GCAAGGCGATGGCTCTCTGGGACACATTTGCTCTATAAGCTCTTTACAAAAGCTTT 184

QY 799 GGTGTTACAGCTCTGATCTTAGAGAGCTATAGATAGCGATTCAAGATAAGTCAT 858
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 185 GGTGGTTTGGCTGATGTTGTTGGCTGCAATAGATCAGGCAGCTGAAGATAATGTTGAC 244
QY 859 GTTCTTTCGCTTTCCCTT 876
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 ATTATCAGCTTGTCCATT 262
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2

Query Match 1.9%; Score 44.8; DB 4; Length 1236;
Best Local Similarity 54.2%; Pred. No. 0.00048;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1614 TGATTCAAGAGAGATTAACTTCACTGTAAATGTCAGGAATCTTCAATGTTGTCCACATGT 1673
Db 596 TCAACCAATTAATGACTATTACACAGCAGCTCTCTGGACATCAATGGCACTCTCCACGT 755

QY 1674 TAGCGGAATCACTGCTCTTATCCGGTCTGCATACCCGAACTGTTCTCCAGCTGCAATCAA 1733
Db 756 AGCTGGTATTGCGCCCTCTTGTCCAAAGCACACCCGAGCTGCACTCCAGCAAGTAAA 815

QY 1734 ATCCGATTCATGACACACAGCGGATTTGTACGATCGTCAGGGAAGC 1781
Db 816 AACAGCCCTCATAGAACTCTGTATATCGTAAAGCCAGATGAATAATAGC 863

RESULT 4
US-08-894-818B-4
; Sequence 4, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/894,818B
  FILING DATE: 20-MAY-1998
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US/08/894,818B
    FILING DATE: 20-MAY-1998
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/JP96/03253
    FILING DATE: 07-NOV-1996
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 323285/1995
    FILING DATE: 12-DEC-1995
  ATTORNEY/AGENT INFORMATION:
    NAME: Browdy, Roger L.
    REGISTRATION NUMBER: 25,618
    REFERENCE/DOCKET NUMBER: TAKAKURA=1
    TELEPHONE: (202) 628-5197
    TELEFAX: (202) 737-3528
    INFORMATION FOR SEQ ID NO: 4:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 1566 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
      MOLECULE TYPE: genomic DNA
      US-08-894-818B-4

Query Match      1.9%; Score 44.8; DB 3; Length 1566;
Best Local Similarity 54.2%; Pred. No. 0.00058;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

US-08-894-818B-4
1614 TGATTCGAAGAAGAGTTAACTTCACTGTAATGTGACGAACTTCAATGCTTTGTCCACATGT 1673
696 TCACCAATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCTCAGT 755
1674 TAGCGGAATCACTGCTTTATCCGGTCTGCATACCCGAACTGGTCTCCAGCTGCAATCAA 1733
756 AGCTGGTATTGACGCCCTCTTGTCTCCAAAGCACACCCGAGCTGGACTCCAGACAAAGTAAA 815
1734 ATCCGGAATGATGACACAGCGGATTTGTACGATCGTCAAGGGAAGC 1781
816 AACAGCCCTCATAGAACTCTGATATCGTAAAGCCAGATGAATAGC 863

RESULT 5
US-08-894-818B-34
Sequence 34, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
  APPLICANT: TAKAKURA, Hikaru
  APPLICANT: MORISHITA, Mio
  APPLICANT: YAMAMOTO, Katsuhiko
  APPLICANT: MITTA, Masanori
  APPLICANT: ASADA, Kiyozo
  APPLICANT: TSUNASAWA, Susumu
  APPLICANT: KATO, Ikunoshin
  TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
  NUMBER OF SEQUENCES: 42
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Browdy and Neimark
    STREET: 419 Seventh Street N.W., Ste. 300
    CITY: Washington
    STATE: D.C.
    COUNTRY: United States of America
    ZIP: 20004
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/894,818B
  FILING DATE: 20-MAY-1998
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/JP96/03253
    FILING DATE: 07-NOV-1996
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 323285/1995
    FILING DATE: 12-DEC-1995
  ATTORNEY/AGENT INFORMATION:
    NAME: Browdy, Roger L.
    REGISTRATION NUMBER: 25,618
    REFERENCE/DOCKET NUMBER: TAKAKURA=1
    TELEPHONE: (202) 628-5197
    TELEFAX: (202) 737-3528
    INFORMATION FOR SEQ ID NO: 3:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 1962 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
      MOLECULE TYPE: genomic DNA
      US-08-894-818B-34

Query Match      1.9%; Score 44.8; DB 3; Length 1962;
Best Local Similarity 54.2%; Pred. No. 0.00068;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

US-08-894-818B-34
1614 TGATTCGAAGAAGAGTTAACTTCACTGTAATGTGACGAACTTCAATGCTTTGTCCACATGT 1673
1092 TCACCAATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCTCAGT 1151
1674 TAGCGGAATCACTGCTTTATCCGGTCTGCATACCCGAACTGGTCTCCAGCTGCAATCAA 1733
1152 AGCTGGTATTGACGCCCTCTTGTCTCCAAAGCACACCCGAGCTGGACTCCAGACAAAGTAAA 1211
1734 ATCCGGAATGATGACACAGCGGATTTGTACGATCGTCAAGGGAAGC 1781
1212 AACAGCCCTCATAGAACTCTGATATCGTAAAGCCAGATGAATAGC 1259

RESULT 6
US-09-445-472-15
Sequence 15, Application US/09445472
Patent No. 6358726
GENERAL INFORMATION:
  APPLICANT: TAKAKURA, Hikaru
  APPLICANT: MORISHITA, Mio
  APPLICANT: SHIMOJO, Tomoko
  APPLICANT: ASADA, Kiyozo
  APPLICANT: KATO, Ikunoshin
  TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
  FILE REFERENCE: TAKAKURA=6
  CURRENT APPLICATION NUMBER: US/09/445,472
  CURRENT FILING DATE: 1999-12-06
  PRIOR APPLICATION NUMBER: 151969/1997
  PRIOR FILING DATE: 1997-06-10
  NUMBER OF SEQ ID NOS: 33
  SOFTWARE: PatentIn version 3.0
  SEQ ID NO 15
  LENGTH: 1962
  TYPE: DNA
  ORGANISM: Artificial Sequence
  FEATURE:
    OTHER INFORMATION: Synthetic
  US-09-445-472-15

Query Match      1.9%; Score 44.8; DB 4; Length 1962;
Best Local Similarity 54.2%; Pred. No. 0.00068;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```

QY 1614 TGATCAAGAGAGTAACTTCACTGTAATGTCAGGAACCTCAATGCTTCTCCACATGT 1673
Db 1092 TCAACCAATTAATGACTATTACACAGAGCTCTCTGGACATCAATGGCAACTCTCAGT 1151
QY 1674 TAGCGGAATCACTGCTCTTATCCGGTCTGCATACCCGAACCTGGTCTCCAGTGCATCAA 1733
Db 1152 ACTGGTATGAGCCCTCTTCTCCAGCACACCCGAGCTGGATCCAGACAAAGTAA 1211
QY 1734 ATCCGATTGATGACAAAGCGGATTTGATACGATCGTCAAGGGAAGC 1781
Db 1212 AACAGCCCTCATAGAAAGCTGTGATGCTAAAGCCGATGAATAGC 1259

RESULT 7

US-09-313-294A-7407
; Sequence 7407, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 7407
; LENGTH: 295
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6476212 700381679H1
; NAME/KEY: unsure
; LOCATION: 4, 282
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-7407

Query Match 1.9%; Score 44.4; DB 4; Length 295;
Best Local Similarity 59.5%; Pred. No. 0.0023;
Matches 75; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 1633 TTCACGTAACTCAGGAACCTCAATGCTTGTCCACATGTTAGCGGAATCACTGCTT 1692
Db 121 TTTCGAATGTTTCTGGAAGTACGATGGTGGCGCATATTGCTGGCATTGCGACTG 180
QY 1693 ATCCGCTCTGCATACCGGAACCTGCTCCAGCTGCAATCAATCCGATTGATGACACA 1752
Db 181 ATAAAGCAGAGAAGCCAAAGTGGAGCCCTCAGCAATAAAATCTGCCCTGATGACACG 240
QY 1753 GCGGAT 1758
Db 241 GCGAAT 246

RESULT 8

US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491

; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc.feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.6%; Score 37.2; DB 3; Length 289;
Best Local Similarity 4.7%; Pred. No. 0.053;
Matches 11; Conservative 105; Mismatches 120; Indels 0; Gaps 0;
QY 1093 GGAGAGTCTTGTATCCGGGAAAGGTATAAAGATCCGGGAGAGAGGTTGAGGTGATT 1152
Db 4 GGRACRPARURURARUR 63
QY 1153 TAGCTCACAGGAGGATAAAGAGTGTCTCTTTGAGAGGCTCACTTCCAAAGAA 1212
Db 64 NNRSNNRNSRNNRNSRNNRNSRNNRNSRNNRNSRNNRNSRNNRNSR 123
QY 1213 GAAATCCGAGCAAAATGTGATTGTCATCGGAGTCAATGGAAGATCGGAGAAAGGA 1272
Db 124 NNRSNNRNSRNNRNSRNNRNSRNNRNSRNNRNSRNNRNSRNNRNSR 183
QY 1273 GAAGCGTTAAAGAGCTGGAGAGCTTGCAATGATCTTAGCCCAATACAGATCAA 1328
Db 184 NNRSNNRNSRNNRNSRNNRNSRNNRNSRNNRNSRNNRNSRNNRNSR 239

RESULT 9

US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc.feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 1.6%; Score 37.2; DB 3; Length 289;
Best Local Similarity 4.7%; Pred. No. 0.053;
Matches 11; Conservative 105; Mismatches 120; Indels 0; Gaps 0;

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	481.8	20.7	2265	10	US-09-938-842A-1065	Sequence 1065, Ap	
2	481.8	20.7	2265	12	US-09-938-842A-1065	Sequence 1065, Ap	
3	398.2	17.1	2295	10	US-09-938-842A-1186	Sequence 1186, Ap	
4	398.2	17.1	2295	12	US-09-938-842A-1186	Sequence 1186, Ap	
5	283.4	12.2	2043	12	US-10-260-238-1206	Sequence 1206, Ap	
6	211.6	9.1	2073	12	US-10-260-238-485	Sequence 485, App	
7	174.6	7.5	740	9	US-09-770-419-59	Sequence 59, Appl	
8	153.2	6.6	606	12	US-10-260-238-5684	Sequence 5684, Ap	
9	141.4	6.1	1467	12	US-10-260-238-703	Sequence 703, App	
10	121.2	4.4	534	12	US-10-260-238-1133	Sequence 1133, Ap	
11	103.2	4.4	510	11	US-09-770-351-228	Sequence 228, App	
12	100.2	4.3	3958	10	US-09-938-842A-338	Sequence 338, App	
13	100.2	4.3	3958	12	US-09-938-842A-338	Sequence 338, App	
14	98.4	4.2	849	12	US-10-260-238-3023	Sequence 3023, Ap	
15	92.4	4.0	828	12	US-10-260-238-3032	Sequence 3032, Ap	

500 AATGGAAGGGAATTTGCCAAGAAAGGAAAGTTTCAGTTCTTCGAGCTGTAAACCGAAGC 559
Db |||||
464 AATGGAAGGGAATGTGAATCTGGTCCGATTTCCGATTTCCAAAGTTGTGTAAACGAAGC 523
Qy |||||
560 TAATCGGTGTAGATTCTTTCATCAGAGGACACCGTGTGCTGTAAATTCACAGAGGAATCAC 619
Db |||||
524 TTATCGGAGGTAGAAGCTTCTCCAAAGGATTTCAAATGGCTTCTGTGTGTGTTTTCGA 583
Qy |||||
620 CAAACATGCTCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCG 679
Db |||||
584 GTAA-----GCGTGAATCTGTTTCTCTCGTGTATGTTGACGGACATGGAACACATATT 637
Qy |||||
680 CCTCAACAGTTGGTGGTCTCTGTTTCGATGCGGAATGTTTCGCAATGAGAGTGGTG 739
Db |||||
638 CAACCTACCGCGCGGGATCCGCGTTAGAAACGCTAGCTTCCCTCGGTACCGCGCGGTA 697
Qy |||||
740 TGGCTCGTGGATGCTCCTGGAGCTCACATTCGAGTCTATAAAGTCTGTGTTCAATG 799
Db |||||
698 CGGCAGAGGTATGGCCACTCGTCTGCTGTTGCTACTTATAAAGTTGTGTGAGTACTG 757
Qy |||||
800 GTTGTTACAGCTGACATTTAGCAGCTATAGATGAGGATTCAGATATAAGTAAAGTCGATG 859
Db |||||
758 GTTGTGTTGGAATCTGATATACTAGTCTGATGATCGAGCTATACCTGATGTTGTTGATG 817
Qy |||||
860 TTCTTTCCGCTTCCCTTGGCGGTTTCCCTATTCTTTGATGATGACACAATCGCCATTG 919
Db |||||
818 TGCTTTCTGTTATCTCTTGGTGGTTCGCTCGGATTTATCCGATACGATTCGATG 877
Qy |||||
920 GAACATTCGAGGCAATCGAAGCGGGTATATCTGTAAATCTGTGAGCTGTGTAAACAGGTC 979
Db |||||
878 GAGCGTTTTCGGCTATCGAGAGAGGTGTTTTCGTTGCTGCTGCTGTTAAATGTTGTTGCTC 937
Qy |||||
980 CAATCGAAAGCTCTGTTGCAACACAGCTCTTGGGTCACACCATTTGGCGAGGACCGC 1039
Db |||||
938 CTACTAGAGTTCTGTTSCCAATGTTGCTCTTGGGTTATGACTGTGTGTTGTTGTTACT 997
Qy |||||
1040 TTGATCAAGATTTCCCGCTGTGTGATCAGATTAGCAACGGAAGCTCTCTATGAGAGT 1099
Db |||||
998 TAGATAGAGTTTTCGGCTTTTCGGAATCTCGTAAACGGAACGACTTACCGGTGTT 1057
Qy |||||
1100 CATTGTATCCGGGAAAGGTATAAGAAATCCGGGAGAGAGTTGAGGTGATTTACGTCA 1159
Db |||||
1058 CGCTGTATAGCGGTAGGAATGGGA-----CGAAGCGCTTGAATGGTTTAAATA 1111
Qy |||||
1160 CAGGAGAGATAAGGAAGTGAATCTGTTGAGAGGTCACATCCAGAGAGAATAATCC 1219
Db |||||
1112 AAGGGAATAGTAGTTCGAGTAACTTTGTTTACCTGCTTCGCTTGATTCGAGTATTGTTTC 1171
Qy |||||
1220 GAGCAAAATGTTGATTTGTATCGGAGTCAATGGAAGATCGGAGAGGAGAGCG 1279
Db |||||
1172 GTGGGAAGATTGTTTGTATAGAGGTGTTAATGCTAGAGTTGAGAAAGGAGCTGTGG 1231
Qy |||||
1280 TTAAGAAGCTGGAGGAGTTCGAATGATCTTAGCCAATACAGAGATCAACCAAGAAGAAG 1339
Db |||||
1232 TTAGAGATGCTGTGTTTGGGATGATAATGGGATATCTGTCGAGTGGAGAGGAGC 1291
Qy |||||
1340 ATTCTATTGAGCTTCATCTCTTACAGCTACATTTGTTTACCTGAGTCACTGAGTCCCTTC 1399
Db |||||
1292 TTGTGGCGGATAGTCAITTTGTTCCCGCATCGCTGTAGGGAAGAGACTGTTGATTTAC 1351
Qy |||||
1400 TGAAGCTTATGTTAATGCCAGCTGAACCAAGGCGCGGATAATTTTGGTGGTACCG 1459
Db |||||
1352 TTAGGAGATGTTTAAAGTCAAGATTTCTAAACCAACCGCTCTTCTGTTTAAAGGAAACGG 1411
Qy |||||
1460 TGATTGGAGGTACAGAGCAGGAGGTGGCTCAGTTTTCAGTTCGAGGACCGAGTTTAG 1519
Db |||||
1412 TCTTGACGTTAAGCGCTCTCTGTTGCTGCTTTTACCTCGAGAGGTCTTAATAGT 1471
Qy |||||
1520 CCAATCTCTCGATCTAAACCGGATATGTTCTCGGAGTCAATATCATTTCCGGCTT 1579
Db |||||
1472 TTACTCTCTGAATCTTGAAGCTGATGTTATTTGCTCTGGAGTTAATTTTGGCTGCTT 1531

1580 GGCCTCAAAATCTAGGACCAACCGGACTTCTTTATGATTTCAAGAAGATTAACTTCACTG 1639
Db |||||
1532 GGTCTGACGCTATTGGTCTTACTGGTCTTGAACAGGACTCTAGAGGACTCAGTTCAACA 1591
Qy |||||
1640 TAATGTGAGGAACCTTCAATGTCTTGTCCACATGTTAGCGGAATCACTGTCTTTATCCGGT 1599
Db |||||
1592 TCATGTGAGTACGTCAATGTCTGCCCACACATCAGTGGTTTAGCGGGTCTTTTGAAG 1651
Qy |||||
1700 CTGCATACCCGAACCTGCTTCCAGCTGCAATCAAAATCCGATTTGATGACAAACAGCGATT 1759
Db |||||
1652 CAGCTCACCTGAGTGGAGTCCGAGTCTATCAAAATCAGCTCTCATGACTACAGCTTACG 1711
Qy |||||
1760 TGTAGATGCTCAAGGAAAGCGATAAAGGATGGT-----AACAAACCGACCGGTTGT 1813
Db |||||
1712 TTCTTGAACAACACCAACGCTCTCTCCATGATGCTGCAGACAACAGCTTATCTAACCCAT 1771
Qy |||||
1814 TTGGGATTTGAGCAGGCGCATGTGAATCGCAAAAGGCGATAAACCCCGGATTTGTTTACA 1873
Db |||||
1772 ATGTGACGGCTCGGGCCATGTAGATCCCAAAAGGCTCTCTCACCAGGCTTGTCTTACG 1831
Qy |||||
1874 ACATTCAACAGTGGATTACATTAATTTACCTCTGCACTCTTGGATTCACAAGATCAGATA 1933
Db |||||
1832 ACATCTCAACCGGAATACATCAGGTTTTTGTCTCTCTAGACTACACAGTCGATCACA 1891
Qy |||||
1934 TTTTAGCAATCACTATAAGAACGTGAGCTGCAATGGAATATTGCGGAAAAAACCAGGTT 1993
Db |||||
1892 TTGTTGGCAT-----TGTGAAGCGACTAGCGTTAACTGCTCGAAGAGTTCTCAGATC 1945
Qy |||||
1994 TTACTCTCAATTACCCGTCGATAGCCGTGATTTTCAACGTCGCAAGACTACGGAGATGA 2053
Db |||||
1946 CTGTGACGCTCAACTACCCCAAGTTTCTCGGTTTTTGGGGGTAAGAAGTTGTGCGGT 2005
Qy |||||
2054 TCACAAGCGGTGTCACATAACGTTGGGAGTCTTAACCTCGATATACTCACTGAATGTCAGG 2113
Db |||||
2006 ACATCGGGAAGTAAACAAATGTTGTCGACGACGCTCGGTTTACAAGTGAAGTTAATG 2065
Qy |||||
2114 CTCAGAGGGGATCAAAAGTTAATTCATCAATCTTAAGACACTGTGTTTCAACACGTCGATC 2173
Db |||||
2066 GAGCTCTAGTGTGCGAATCTCTGTTAAACCATCGAAACTTTCGTTTAAAAAGCTGGAG 2125
Qy |||||
2174 AGAGCTGAGCTATAGAGTATGTTTGTATTGAAGAAGAAAAACAGAGGAGGAGGTGG 2233
Db |||||
2126 AGAAGAAGAGGTACACAGTCACTGTTTGTAGCAAGAAGAGGTAGTATGACGAACAAGG 2185
Qy |||||
2234 CTAGCTTGG 2242
Db |||||
2186 CTGAGTTTG 2194

RESULT 2

US-09-938-842A-1065
; Sequence 1065, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepes, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1065
; LENGTH: 2265
; TYPE: DNA

; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1065

Query Match 20.7%; Score 481.8; DB 12; Length 2265;

Best Local Similarity 55.3%; Pred. No. 1e-141;

Matches 1033; Conservative 0; Mismatches 812; Indels 24; Gaps 4;

```
QY 380 ACTCCGCTGATGGTCTAAATCTCGTTTGGTCAAGGCACAAATATCGCGGTGCTTGATA 439
DB 344 AATTCGGTGTTCAGATCTCGGTCTCTTCTTAACGGCGTTATCATCGGAGTTTATAGATA 403
QY 440 CTGGAGTTTGGCCTGAAGTCTTAGCTTTCAGCATACCGAATGCCCTTGATCCACGGA 499
DB 404 CTGGCGTTTGGCCTGAATCTAGAGCTTCGATGATCTGATATGCTGATGATCTTCTCTA 463
QY 500 AATCGAAGGATTTGGCAAGAGGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGC 559
DB 464 AATGGAAGGAGATGTGAATCTGTTCCGATTTCCGATTTCCGATTTCCGATTTCCGAT 523
QY 560 TAATCGGTGTAGATTTCTCATCAGAGGACACCGTGTGCTTAATTCACAGAGGAATCAC 619
DB 524 TTATCGGAGCTAGAAAGCTTCTCAAGAGATTTCAAAATGGCTTCTGGTGGTGTGTTTCGA 583
QY 620 CAACATGCTCGGTGATATATTTCCGAGAGATTTCAACGGGACACGGGACTCACACCG 679
DB 584 GTAA-----GCGTGAATCTGTTTCTCTCGTGTATGACGACATGGAACACATCTT 637
QY 680 CCTCAACAGTGTGGTGGTCTCTGTTTCGATGGCAATGTTCTTGGCAATGGAGCTGGTG 739
DB 638 CAATACCGCGGGGATCCCGGTTAGAAACGCTAGCTTCTTCGGTTACCGCCCGGTA 697
QY 740 TGGCTCGTGGATGGCTCTCGAGCTCACATTCAGTCTATAAAGTCTGTGGTTCGAATG 799
DB 698 CGGCCAGAGTATGGCCACTCGTCTGCTGCTGCTACTTATAAAGTTTGTGGAGTACG 757
QY 800 GTTGTTACAGCTCTGACATTTCTAGCAGCTATAGATGAGCATCAAGATAAAGTGCATG 859
DB 758 GTTCTTTTGGATCTGATATATAGTCTGCTATGATCGAGCTATCTTGTGTTGTTGATG 817
QY 860 TTCTTTTGGCTTCCCTTGGGTTTCCCTATTCTTTGATGATGACAAATCGCCATTG 919
DB 818 TGCTTTTGGTATCTCTTGGTGGTGTCTGCTCGGTATATCGGATAGCATTTGCCATTG 877
QY 920 GAACATTCGAGCCATGAAACCGGTATATCTGTAATCTGTCAGCTGGTAAACACGGTTC 979
DB 878 GAGGTTTTCGGCTATCGAGAGAGTGTGTTTGTGCTGTTCTGCTGTTGATAGTGTGTC 937
QY 980 CAATCGAAGCTCTGTTGCAACACAGCTCTTGGGTCTCAACCATTTGGCGCAGGCAAGC 1039
DB 938 CTACTAGAGCTTCTGTTGCCAATGTTGCTCTTGGGTTATGACTGTTGTTGCTGTTACT 997
QY 1040 TTGATCGAAGATTTCCGCTGTGTGATAGATAGCAACGGAAGCTTCTCTATGGAGAGT 1099
DB 998 TAGATAGAGATTTTTCGGCTTTTTCGGAATCTCGTAAACGGGAAACGACTTACCGGTG 1057
QY 1100 CATTTGATCCGGGAAAGGTATAAGATCCCGGAGAGAGGTTGAGGTGATTTACGTCA 1159
DB 1058 CGCTGTATAGCGGTAGGAATGGGA-----CGAAGCCGCTTGAATTTGTTTATAATA 1111
QY 1160 CAGGAGGAGATTAAGGAAGTGAATTTGTTTGAAGGGTCACTTCCAAAGAGAGAATATCC 1219
DB 1112 AAGGGAATAGTATTCAGTAATCTTTGTTTACTGTTCTGTTGATTCAGTATTGTTTC 1171
QY 1220 GAGCAAAATGGTATTTGATCGCGAGTCAATGGAAGATCGGAAAGGAGAAAGCGG 1279
DB 1172 GTGGGAAGATTTGTTTGTGATAGAGTGTATATGCTAGAGTTGAGAAAGGAGCTGTGG 1231
QY 1280 TTAAGAGAGCTGGAGGAGTTGCAATGATCTTAGCCATACAGAGATCAACCAAGAGAAG 1339
DB 1232 TTAGAGATGCTGGTGGTTTGGGATGATAATGCGAATACTGCTGCGAGTGGAGAGAGAC 1291
QY 1340 ATTCTATTGAGTTCACTCTCTTACCAGCTACATTGTTGTTACATGATGAGTCACTCTTC 1399
```

```
DB 1292 TTGTGGCGGATAGTCAATTTGCTTCCCGGATCGCTGTAGGGAAGAGACTGGTGATTAC 1351
QY 1400 TGAAGGCTTATGTTAATGCGCACCGGTGAAAACAAAGGCGCGGATAAATTTTGGTGGTAGCG 1459
DB 1352 TTAGGAGTATGTTAAGTCAAGTCTAAACCAACCGCTCTTCTTGTGTTTAAAGAAACGG 1411
QY 1460 TGATTTGGAGGTCAAGGACCGGAGGTGGCTAGTTTTCAGCTCGAGGACCGAGTTTAG 1519
DB 1412 TTTCTTGAAGTTAAGCGGTCTCTGTTGTTGCTTTTAGCTCGAGAGGTCTTAATACTG 1471
QY 1520 CCAATCCCTTCGATATAAACCAGGATATGTTGCTCCGGGAGTCAATATCATTCGCGGTT 1579
DB 1472 TTACTCTGAAATCTTGAAGCTGATGTTATTTGCTCTGGAGTTAATATTTTGGCTGGTT 1531
QY 1580 GGCCTCAAAATCTTAGGACCAACCGGACTTCCTTTATGATTCAGAGAGTTAACTTCACTG 1639
DB 1532 GGTCTGACGCTATTGGTCTTACTGCTTGAACAGGACTCTAGGAGGACTCAGTTCAACA 1591
QY 1640 TAAATGTCAGGAACCTCAATGCTTTGTCCACATGTTAGGGGAATCAGTCTCTTATCCGGT 1699
DB 1592 TCATGTCAGGTACGTCATATGTCATGCCACACATCAGTGGTTTAGCGGCTCTTTTGAAG 1651
QY 1700 CTGCATACCCGAATCTGCTCCAGCTGCAATCAAAATCCGCAATTCATGACACACGCGATT 1759
DB 1652 CAGCTCACCTGAGTGGAGTCCGAGTCTATCAAAATCAGCTCTCATGACTACAGCTTACG 1711
QY 1760 TGTACGATCGTCAAGGGAACCGGATAGGATGGT-----AACAAACCCAGCGGTGTG 1813
DB 1712 TTTTGAACAACCAACGCTCTCTCCATGATGCTGCAGACAACAGCCTATCTAACCCAT 1771
QY 1814 TTGGGATTTGGAGCAGGCGATGTGAATTCGGAAGGAGGATAAACCCGGGATTTGTTTACA 1873
DB 1772 ATGCTCAGGCTCGGCGCATGTAGATCCCAAAAGGCTCTCTCACCAGGCTTGTGTACG 1831
QY 1874 ACATTCACCAAGTGGATTAAGTAACTTACCTCTGCACTCTTGGATTCAACAGATCAGATA 1933
DB 1832 ACATCTCAACCGAGGAATACATCAGGTTTTTGTCTCTCTAGACTACACAGTCGATCACA 1891
QY 1934 TTTTAGCAATCACTATAAGAACGTGAGCTGCAATGGAATATTGCGGAAAACCCGGGTT 1993
DB 1892 TTGTTGCAAT-----TGTGAAGGACCTAGCTTAACTGCTCGAAGAGTTCTCAGATC 1945
QY 1994 TTAGTCTCAATTAACCGCTCGATAGCCGTGATTTTCAACCGTGGCAAGACTACGAGATGA 2053
DB 1946 CTGGTCACTCAACTACCCAAAGTTTCTCGGTTTTTGTGGGGGTAAAGAGTTGTGCGGT 2005
QY 2054 TCACAGCGGTGTCACTAACGTTGGGAGTCTTAATCTGATATACTCAGTGAATGTCAAGG 2113
DB 2006 ACATTCGGGAAGTAAACAAATGTTGGTGCAGCAAGCTCGGTTTACAAAGTGCAGGTTAATG 2065
QY 2114 CTCCAGAGGGGATCAAGTATTGTTCAATCTTAAGAGACTTGTCTTCAACACGTTGGATC 2173
DB 2056 GAGCTCTCTAGTGTGGGAATCTCTGTAAACCATCGAAACTTTTCGTTTAAAGCGTGGAG 2125
QY 2174 AGACCTCAGCTATAGATGATGTTGTTGATTTGAAGAGAGAAAAACAGAGGAGGAGTGG 2233
DB 2126 AGAAGAGAGGTACACAGTACAGTTGTTGTACAGAGAGGAGTGTAGTATGACGACAGG 2185
QY 2234 CTAGCTTTG 2242
DB 2186 CTGAGTTTG 2194
```

RESULT 3

US-09-938-842A-1186
; Sequence 1186, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

;; TITLE OF INVENTION: SAME, AND METHODS OF USE
;; FILE REFERENCE: SRIPI300-3
;; CURRENT APPLICATION NUMBER: US/09/938,842A
;; CURRENT FILING DATE: 2001-08-24
;; PRIOR APPLICATION NUMBER: US 60/227,866
;; PRIOR FILING DATE: 2000-08-24
;; PRIOR APPLICATION NUMBER: US 60/264,647
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/300,111
;; PRIOR FILING DATE: 2001-06-22
;; NUMBER OF SEQ ID NOS: 5379
;; SEQ ID NO 1186
;; LENGTH: 2295
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1186

Query Match 17.1%; Score 398.2; DB 10; Length 2295;

Best Local Similarity 53.2%; Pred. No. 4.5e-115;

Matches 1072; Conservative 0; Mismatches 898; Indels 45; Gaps 9;

QY	205	TCTCGACTTCTCTACTCTGCTCTGCGATTGAGGATTGCTGCTCAGTTGACTGAA	264
DB	166	TCTCGAATCGCCATGTTTACCACACAGTCTTCCAGTGTCTCCCGCGTTGTACTCCA	225
QY	265	TCAGAACCCGAGATACCTGAGATATTCACCTGAAAGTTGTTGCAGTGAGACCTGACCATGTT	324
DB	226	GATGAACGAGATATCTCCGTACCCAGCCAGAGTCTTGTGCTTTTCGNAAGCCAGCT	285
QY	325	CTTCAGGTTCAAAACCACTTACTCTTACAAAGTTCTTGCGACTCGACGGTTTGGAAACTCC	384
DB	286	CGAGAGCTTCACACCACACGTTCTCCTCAATTTCTTGGTTTACAA-----AACCAAAA	339
QY	385	GGTGATATGTTCTAAATCTCGTTGTCAGGACACAAATATCGCGTGTCTGATCTGGA	444
DB	340	GGACTATGTCAGAACTGATTAACGATCAGACGTAATCATTTGGCGTTTTCACACCGGA	399
QY	445	GTCTGGCTGAAAGTCTAGCTTTGACGATACCGGAATGCTTCGATTTCACGGAATGG	504
DB	400	ATTTGGCCGAGCGAGAGATTCTCAGATCTTAACTCGTCCCAATTCGAAAGGTGG	459
QY	505	AAAGGATTTGCCAAGAGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGTAAATC	564
DB	460	AGAGCGTTTCGGAATCCGGAGCCAGATTTCAGTCTCGGAATGTAAACGTTAAATTC	519
QY	565	GGTCTAGATTCTTCATCAGAGACACCGTGTGCTTAATTCACAGAGGAATCACCAAC	624
DB	520	GGAGCAAGATTCTCGCTAAGGGACAAACAGCGCT--GTAATCGGAGGAATCAACAAA-	576
QY	625	ATGCTCTGATACATTTCCGCAAGAGATTCAACGGGACACGGACTCACACCGCTCA	684
DB	577	---ACCGTTGAGTTTCTATCTCTCGTGAACCGGATGGACACCGTACTCACCTTCTCA	633
QY	685	ACAGTTGGTGGATCCTCTGTTTCGATCGGGAATGTTCTTGGCAATGAGCTGTGTGGCT	744
DB	634	ACCGCGCTGGCGGTACGCTTTTAAAGCAGATATGTCGGTTACGCTCCGCTGTAGCC	693
QY	745	CGTGGATGCTCTCGAGCTCAATTCGAGTCTATAAAGTCTGTGGTTCAAT---GGT	801
DB	694	AAAGTGTGTCTCCAAAAGCTCGTATCGCGCTACAAAGTCTGTGGAAAGATTCGGT	753
QY	802	TGTTACAGCTCTGACATCTTAGCAGCTATAGATGACGATTCAGATAAGTCGATGTT	861
DB	754	TGCTCTGATCCGATATCTCGCGCTTTGATGCGCGCTTTGATGCGCGCTTTAGAGACGGT	813
QY	862	CTTTTCGGTTTCCCTTGGCGTT-----TCCTATTCTTTGATGATGACAAATC	912
DB	814	ATATCGATCTCAATCGGTGTGGAGACGGGATTAATTCGCGGTATTAACCTCGATCCAATC	873
QY	913	GCCATTGGACATTCGAGCCATCGAAGCGGCTATATCTGTAATCTGTGAGCTGTTAAC	972
DB	874	GCTATAGCTGTACGGCGCGCGCTCGAAAGGAATCTTCGTCTCTTCTCTCGCGGAAC	933

QY	973	AACGGTCCAATCGAAGCTCTGTTGCANACACAGCTCCTTGGGTCTCAACCATTTGGCGCA	1032
DB	934	GAAGGACCTAACCGTATGTTCAGTATCAGAACCTCGCGCGTGGTAAACACCGTTGGTGT	993
QY	1033	GGACCGCTTGATCGAAGATTTCCCGCTGTGGTCAGATTAGCCAAACGGAAGGCTTCTCTAT	1092
DB	994	AGTACAAATCGATCGGAATTTCCAGCGGATGCTATTCTCGCGACGACATCTCTCAGA	1053
QY	1093	GGAGAGTCATTGTATCCGGGAAAAGGTATAAGAAATCCCGGAGAGAGGTTGAGGTGATT	1152
DB	1054	GGAGTGTCTCTTTACGCTGGAGTACCTTTAAACGGTGTATGTTTCCGGT---GGTTTAT	1110
QY	1153	TACGTACAGAGGAGATAAAGAGTAGTGTCTGTTTGAGAGGTCACCTTCCAAGAGAA	1212
DB	1111	CCCGTAAATCGGGAATGTCTATCAGCGCTCTCTATGTATGAGAAACACGCTTGATCCGAAG	1170
QY	1213	GAAATCCGAGCAAAATGGTGATTTGTATCGGAGTCAATGGAAGATCGGAAAGGA	1272
DB	1171	CAAGTAGGGGTAATAATAGTAATCTGCGATAGGAAGACAGTCCACCGTAGCCAAAGGA	1230
QY	1273	GAAGCGGTTTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCCCAATACAGAGATCAACCAA	1332
DB	1231	TTGGTTGTGAAGAAAGCAGGTGGTGTGCGAATGATCTCGCTAATGGAGCATCTAACGGT	1290
QY	1333	GAAGAAGATCTATTGACGTTCTCTTACAGCTACATTTGATTTGTTTACATGAGTCA	1392
DB	1291	GAAGGATTAGTCGAGATGCTCATCTTATTCAGCCTGTGCCCTTGGATCAACAGGAAGA	1350
QY	1393	GTCTCTTGAGGCTTATGTTAATGCCAGGTGAAACCAAGCGCGGATAATTTTGGT	1452
DB	1351	GATGAATCAAGCATATGCTTCTTACATCCGAATCCAATTTGCTTCAATTTGATTTTCTCAGA	1410
QY	1453	GGTACGCTGATTGGAGGTCACGAGCACCGAGGTGCTCAGTTTTCAGCTTCGAGGACCG	1512
DB	1411	GGACTATAGTTGGATTAACCGGCTCCGGTTATGCTTCTTCTCCGGTAGAGGACCA	1470
QY	1513	AGTTTAGCCAACTCTTCGATATAAAACCGGATATGATGCTCCGGAGTCAATCATTT	1572
DB	1471	AACGTTTAAAGCCCGAGATTTCTTAAACCGGATTTGATTTGCTCCCGAGTTAATCCTC	1530
QY	1573	CGCGTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAGAGAGGTTAAC	1632
DB	1531	CGCGATGGACAGACGCTGTTGGACCTACAGGTTTGGCGTCAGATCCAGGAAACCGAA	1590
QY	1633	TTCACTGTAATGTCAGGAATCTTCAATGCTTGTCCACATGTTAGCGGAATCACCTGCTCTT	1692
DB	1591	TTCAACTTCTCTCCGTAATCTTCAATGGCATGCTCTCAGCTAGTGTGCGCGCGCTT	1650
QY	1693	ATCCGCTCTGCATACCCGAACTGCTCTCCAGCTGCAATCAAAATCCGCAATGATGACAAAC	1752
DB	1651	CTCAATCCGCTCATCCAGATTGGAGCCCTGCGGTGATACGATCGGCAATGATGACAAACG	1710
QY	1753	GGGATTTGTACGATTCGTCAGGGAAGCGAATAAGGATGTTAAAC-----AAACGAGCC	1806
DB	1711	ACTAACCTCGTCGATAACTCTAACCGCTCGTTGATCGATGAATCCACCGGAAATCGGCT	1770
QY	1807	GGTGTGTTTCGATTGGAGCAGGCAATGTAATCCGCAAAAGGCGATAAAACCCGGGATTG	1866
DB	1771	ACGCGTTATGATTAACGGGTCCGGTCAATTAATTTGGGCGGGGTATGAATCCCGGCTT	1830
QY	1867	GTTTACAACTTCAACAGTGGATTACATACTTACCTCTGCACTCTTGGATTCAACAGA	1926
DB	1831	GTCTACGATATACTAATGATGATTAATACGTTTCTTGTCTCCTCGGTACGGACCA	1890
QY	1927	TCAGATATTTTAGCAATCATCTATAAGAACTGAGCTGCAATGGAATTTTCGGGAAAC	1986
DB	1891	AAGACGATCCAAAGTGATAACAAAGAACACCGGTGAGATG---TCGAGACACGAGAAACCG	1947
QY	1987	CGCGGTTTTAGTCTCAATTAACCGTCGATACCGGTGATTTC-----AAACGTGGC	2037
DB	1948	TCTCCCGGAAATTTGAATTTATCTTCGATCAGCGCGGTGTTTCTTACTTAATAGAGAGGA	2007
QY	2038	RAGACTACGGAGATGATCACAAGCGGTGTCATTAACGTTGGGAGTCTCTTAACCTCGATATAC	2097

Db 2008 TTGGTGAATAAACTGTTATTAAGGACGGGACGATGTCGGGAGGCTGAGCGGTTAT 2067
 Qy 2098 TCAGTGAATGCAAGCTCCAGAGGGATCAAAATGTTGTCATCTTAAGAGACTGTG 2157
 Db 2068 CGGCGAGGATAGAGTCCGAGAGGAGTACGGTGACAGTGAACCACTAGGCTGTG 2127
 Qy 2158 TTCAACACGTTGATCAGACCGCTGAGCTATAGAT 2192
 Db 2128 TTACTTCGGCGTTAAGAGACGGAGCTATCGGT 2162

RESULT 4

US-09-938-842A-1186
 ; Sequence 1186, Application US/09938842A
 ; Publication No. US20040009476A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; FILE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SCRIPI300-3
 ; CURRENT APPLICATION NUMBER: US/09/938.842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264,647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300,111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 1186
 ; LENGTH: 2295
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-09-938-842A-1186

Query Match 17.1%; Score 398.2; DB 12; Length 2295;
 Best Local Similarity 53.2%; Pred. No. 4.5e-115;
 Matches 1072; Conservative 0; Mismatches 898; Indels 45; Gaps 9;
 Qy 205 TCTCGACTTCTACTCTATGGCTCTGCGATTGAAGGATTGCTGCTCAGTTGACTGAA 264
 Db 166 TCTCGAATCGTCCATGTTTACCACACAGTCTTCCATGTTTCTCGCGGTGTTACTCCA 225
 Qy 265 TCAGAACCGGAGATCTAGATATTCACCTGAAGTTGTCAGTGAGACCTGACCATGTT 324
 Db 226 GATGAAGCAGATAATCTCCGTAACCCACCCAGCAGTCTTCTGTTTTCGAAGACCGAGT 285
 Qy 325 CTTGAGGTTCAACACATTAATCTTACAGTTCTTGGACTCGACGGTTTGGAACTCC 384
 Db 286 CGAGAGCTTCAACACACAGTCTTCTCCTCAATTTCTTGTTTACAA-----AACCAAAA 339
 Qy 385 GGTGTATGCTCTAAATCTCGGTTTGGTCAAGGCAATATTCGGCGTGTGATCTGGA 444
 Db 340 GGACTATGTCAGATCTGATTACGGATCAGACGTATCATCTGGGTTTTCACACCGGA 399
 Qy 445 GTTTGGCTGAAAGTCTTAGCTTTTACGATACCGGAATGCTTCGATTCCACCGAAATGG 504
 Db 400 ATTTGGCCGGAGCGGAGGAGTTTCTCAGATCTTAACCTCGGTCCCAATTCAAAAGGTGG 459
 Qy 505 AAAGGATTTCCGAGAGAGGAGAAAGTTTCAAGTTCTTCGAGCTGTAAACCGGAGCTAATC 564
 Db 460 AGAGGCGTTTCGATCCGGAGCCAGATTTCAGTCTCGGAATGTAACCGGTAAATATATC 519
 Qy 565 GGTGTAGATTTCTCATCAGAGGACACCGTGTCTGCTAATTCACAGAGGATCAACCAAC 624
 Db 520 GGAGCAAGATTTCTCGTTAAGGACAAACACCGCT--GTAAATCGGAGGATCAACAA- 576
 Qy 625 ATGCTCTGTAATCATTTCCGCAAGAGATTCAACGGGACACGGGAGTCTCACACCGCTCA 684

Db 577 ---ACGGTTGAGTTTCTATCTCTCGTGACGCCGATGCACACGGTACTCACACTTCTCTCA 633
 Qy 685 ACAGTTGGTGATCTCTGTTTCCGATGCGGAATGTTCTTGGCAATGGAGCTGGTGTGCT 744
 Db 634 ACCGCCGCTGGCGTACAGCTTTTAAAGCGAGTATGTCGGTTACGCCCTCCGGTGTAGCC 693
 Qy 745 CGTGGATGGCTCCCTGGAGCTCACATTGCACTATAAAGTCTGTTGTTCAAT---GGT 801
 Db 694 AAAGTGTGCTCCAAAGCTGTTATCGCGCTACAAAGTCTGTTGAAAGATTCGGGT 753
 Qy 802 TGTTCAGCTCTGACATTTAGCAGCTATAGATGTAGCGATTCAAGATAAAGATTCGATGTT 861
 Db 754 TGTCTCGATTCCGATATTCTGCGCGCTTTGATGCGCTGTTAGAGACGGTGTGCGAGTT 813
 Qy 862 CTTTCGCTTTCCCTTGGCGGTT-----TCCCTATTCTCTTGTATGATGACACAATC 912
 Db 814 ATATCGATCTCAATCGGTGGTGGAGACGGGATTACTTCCCGCTATTACTCGATCAATC 873
 Qy 913 GCCATTGGAACATTCGAGGCCATGAAACGCGGTATATCTGTAATCTGTCAGCTGGTAAC 972
 Db 874 GCTATAGGCTCGTACCGCGCGCGTGAAGGAATCTTCGTTCTTCTCTGCGCGGAAC 933
 Qy 973 AACGGTCCAAATCGAAAGCTCTGTTGCAACACAGCTCCTTGGGTCTCAACATTCGGCA 1032
 Db 934 GAAGACCTAACGGTATGTGCTAGTACGAACCTCGCGCGTGGTAAACCACTGTTGCTGT 993
 Qy 1033 GGCACGCTTGATCGAAGATTTCCCGCTGTGTCAGATTAGCCAAACGGAAGCTTCTCTAT 1092
 Db 994 AGTACAATCGATCGGAATTTCCAGCGATGCTATTCTCGCGACGGACATCGTCTCAGA 1053
 Qy 1093 GGAGAGTCAATGTTATCCGGGAAAGGTATAAAGATGCGGGAGAGAGGTTGAGTGAAT 1152
 Db 1054 GGAGTGTCTCTTTACGCTGGAGTACCTTTAAACCGTCTGATGTTTCCCGT---GGTTTAT 1110
 Qy 1153 TACGTACAGAGGAGAGATAAAGGAAGTGAAGTCTGTTTGGAGGGTCACTTCCAAAGAGA 1212
 Db 1111 CCGGTAATCGGGAATGTCATCAGCTCTCTATGATGGAGAACACGCTTGTATCGGAAG 1170
 Qy 1213 GAAATCCGAGGCAAAATGTTGATTCGCGAGTCAATGGAGATCGGAGAAAGA 1272
 Db 1171 CAAGTCAGGGTAAATAGTATCTCGATAGAGGAAGCAGTCCACGCTAGCACAAGGA 1230
 Qy 1273 GAAGCGGTTAAAGAGCTGGAGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCA 1332
 Db 1231 TTGGTTGTGAAGAAAGCAGGTGGTGTGGAATGATTCTCGCTAATGGAGCATCTAACGCT 1290
 Qy 1333 GAAGAGATCTTATTGACGTTTCATCTTACGAGTACATGATTGTTGTTTACACTGATCA 1392
 Db 1291 GAAGGATTAGTCGAGATGCTCATCTTATTCAGGCTGTGCGGTTGGATCAACCAAGGA 1350
 Qy 1393 GTCTCTTGAAGGCTTATGTTAATGCCAGGTTGAAACCAAGGCGCGGATATTTTGGT 1452
 Db 1351 GATAGATCAAGACATATGCTTCTTCATCCGATCCAAATTCGCTTCAATGATTTCAGA 1410
 Qy 1453 GGTACGCTGATTGGGAGTCAAGACACCGGAGTGGCTCAGTTTTCAGCTCGAGACCG 1512
 Db 1411 GGAACATATAGTTGGGATTAAACCGGCTCGGTTATTGCTTCTTCTCCGTTAGAGACCA 1470
 Qy 1513 AGTTAGCCAAATCTTCGATATAAAGCGGATGATGTTGCTCCGGAGTCAATATCAT 1572
 Db 1471 AACCGTTTAAAGCCCGGAGATTCTTAAACCGGATTTGATTGCTCCCGGAGTTAACATCCTC 1530
 Qy 1573 GCGGCTTGCCCTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAAGAGAGTTAAC 1632
 Db 1531 GCGCATGACAGACGCTGTTGGACCTACAGGTTTGGCGTCAATCCAAAGGAAACCGAA 1590
 Qy 1633 TTCACTGTAATGTCAGGAATTTCAATGCTTCTCCACATGTTAGCGGAATCACTGCTCT 1692
 Db 1591 TTCAACATTTCTCTCGGTACTTCAATGGCATGTTCTCAGTTAGTGTGTCGCGCGGCT 1650
 Qy 1693 ATCCGCTGTGCATACCGGAATGTTCTCCAGCTGCAATCAATCCGCAATGTATGACAACA 1752
 Db 1651 CTCAAATCCGCTCATCCAGATTGGAGCCCTGCGGTGATACGATCGGAATGATGACAACG 1710

1753 GCGGATTGTACGATCGTCAAGGGAAGCGATAAAGATGGTAAC-----AAACGAGCC 1806
1711 ACTAACCTCGTGATAACTTAACCGCTCGTGTGATCGATGAATCCACCGGGAATCGGCT 1770
1807 GGTGTGTTTGGATTGGAGCAGGCGATGTAATCCGCAAAAGCGGATAAACCAGGATG 1866
1771 ACGCTTATGATTACGGGTGCGGTCAATTAATTTGGCGGCGCTAAGAAATCGGGTCTT 1830
1867 GTTTTAAACATTCAACAGTGGATTACATAAATTAACCTCTGCACTCTTGGATTCAAGA 1926
1831 GTCTACGATATACTAATGATGATTACATTAACGTTCTTGTCCATCGGTAAGCA 1890
1927 TCAGATATTTAGCAATCACTCATAGAACGTCGAGTCGATGCAATGGAATATTTGCGGA 1986
1891 AAGACGATCAAGTGAATAAGAACACACCGGTGAGATG---TCCGACGACCGAGGAACCG 1947
1987 CCGGTTTTAGTCTCAATTAACCGTGCATAGCGGTGATTTTC-----AAACGTGGC 2037
1948 TCTCCCGGGAATTTGAATTAATCTTCCGATCACCGCGGTGTTCTCTACTAATAGAGGA 2007
2038 AAGACTACGAGATGATCAAGCGGTGTCACATAAGCTTGGGAGTCTTAACCTCGATATAC 2097
2008 TTGTGTAGTAAACTGTTTAAAGACCGCGACGAATGTCCGGCAGGCTGAGGCGGTTTAT 2067
2098 TCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTTATTTGTCATCTTAAGAGACTTGTG 2157
2068 CGGCGGAGGATAGAGTCCGCGAGGAGGTGACGCTGACAGTGAACCACTAGGCTTGTG 2127
2158 TTCAACACGTTGGATCAGAGCTGAGCTAGCTATAGAT 2192
2128 TTTACTTTCCGCGGTTAAGACGCGAGCTATGCGGT 2162

RESULT 5

US-10-260-238-1206
; Sequence 1206, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Krebs, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1206
; LENGTH: 2043
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (389)---(389)
; OTHER INFORMATION: n = any nucleotide

Query Match 12.2%; Score 283.4; DB 12; Length 2043;
Best Local Similarity 52.4%; Pred. No. 1.5e-78;

Matches 825; Conservative 0; Mismatches 696; Indels 54; Gaps 7;
QY 647 CAAGAGATTCAACGGGACACGGGACTCACACCCCTCAACAGTTGGTGGATCTCTGTTT 706
DB 446 CAACGGATGATAGTGGCCATGGGACACATACCCCAAGCAGCGAGCTGGCGTTTGTAG 505
QY 707 CGATGGCGAATGTTCTTTGGCAATGGAGCTGCTGGCTCGTGGATGCTCTCTGGAGCTC 766
DB 506 ATGTTGAAGCGTGTGTTGGCAATGGAAATGGCAGCAGCTGGATGGCTCTCTGTCAC 565
QY 767 ACATTGCACTCTAATAAGTCTGTTGGTTCAATGGTTTACAGCTCTGACATCTTAGCAG 826
DB 566 ACCTTGCTATCTACAAGTGTGACGCGATAAAGTTTCCGCTGATCTGACATATACTTCTG 625
QY 827 CTATAGATGATGATCAAGATTAAGTCGATGTTCTTTGCTCTTCCCTTGGCGGTTTCC 886
DB 626 GCATGGAGCGGCTATTTGCTGCTGTTGACATCATGCTATGCTCTCTTTGGTGGTCCAG 685
QY 887 CTATTTCTTTGATGATGACACATCGCATTTGGAACATTCGCGACCATCGAAGCGCGTA 946
DB 686 CGAAACCAATTTTCAATGATATAATAGCCACTGATCTTTTTTCTGCCATGAGGAAGGGA 745
QY 947 TATCTGTAATCTGTGCAGCTGGTAAACACCGTCCAAATCGAAAGCTCTGTTGCAAAACAG 1006
DB 746 TCTTTGTAAGTCTTGTGCGAGAAATTTCTGCTCATCTTCCAGCACTTAAGCAACGAG 805
QY 1007 CTCCTTGGGTCTCAACCATTTGGCGCAGCAGCTTGTGATCGAAGATTTCCGCTGTGTCA 1066
DB 806 CACCATGGTCTCTGCTGTTGGTCCAGCAGATAGATAGACAAATGGAAGCCCTTGTCA 865
QY 1067 GATTAGCCACGGAAGCTTCTCTATGAGAGTCAATGTTATCGGGAAGAGGTATAAGA 1126
DB 866 AGCTAGTGTGAGAGCTTATTTGTTGGCGAATCTGCTTATCAACCAATTAATCTTATC 925
QY 1127 ATGCCGGAGAGAGGTTGAGGTGATTTACGTCAAGGAGGAGATAAAGGAAGTCTTCT 1186
DB 926 CCTTAGATTAGTGTACCCACAGACTTCC-----GGTCAAAATTTATGCTTCT 973
QY 1187 GTTTGAGAGGCTCACTTCCAAAGAGAAGAAATCCGAGGCAAAATGCTGATTTGATCCCG 1246
DB 974 TCTGA-----AAGATGTTGAGGAAGATGTTGCTCTGTGAGCACA 1015
QY 1247 GAGTCAATGGAAGATCGGAGAAAGAGAGCGGTTAAAGAGCTGGAGGATGTCATGA 1306
DB 1016 CA---ACCTCATCAGACATTTATTGAGCGCTTCGTCAAGGATGCTGGTGCATCTGGGTGA 1072
QY 1307 TCTTAGCCAAATACAGAGATCAACCAAGAGAAGATTTCTATTGACGTTTCATCTTACCAG 1366
DB 1073 TCTTCTTAGTACAGAGGATAGTGGCCACATAACATTTGCTGACCCAAATGCTCTACCGG 1132
QY 1367 CTACATTTGATTGGTTTACACTGAGTCAGTCTTCTGAAAGGCTTATGTTAAATGCCAGGTGA 1426
DB 1133 TGTCTTATGTGGATTTCCAGATGCACAGTAATCAGACAGTACATTAATCAAGTAACA 1192
QY 1427 AACCAAAGCGCGGATTAATTTTGGTGTACGGTGTATGGAGGTACAGGACCGGAGG 1486
DB 1193 GTCCAACTGCCTCAATCATCTTCAATGGTACCTCATTTGGGGAAGAACTCAAGCTCCTGTAG 1252
QY 1487 TGGCTCAGTTTTTTCAGCTCGAGACCGAGTTTAGCCATCTTCCGATCTAAACCGGATA 1546
DB 1253 TTGCAATTTTCTCATCTCTGCGCCCTAGTACTGCGAGCCCGAGGTATTTCTTAACTGACA 1312
QY 1547 TGATTGCTCCGGGAGTCAATATCATTTGGGCTTGGCGCTCAAAATCTAGGACCAACCGGAC 1606
DB 1313 TCATTTGGTCTGGAGTAAACGTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1372
QY 1607 TTTCTTATGATTCAAGAAAGATTAACTTCACTGTAATGTACAGGAATTTCAATGTTGTC 1666
DB 1373 ACAATGATAAACAACAGGACATTTCACTGC-----TTATCTGGAACATCAATGTCAACTC 1426
QY 1667 CACATGTTAGCGGAATCATCTGCTTTATCCCGTCTGCAATACCCGAACTGCTCTCCAGTG 1726
DB 1427 CTCACTTAGTGGCATTTGCTGCTCTCATCAAGGGAACACACCCAGATTGGTCTATCCCGCAG 1486

QY 1727 CAATCAAAATCCGATTTGATGACAAACAGCGGATTTGTACGATCTCAAGGAAAGCGATAA 1786
DB 1487 CGATCAAGTCCGCAATCATGACACAGCTTATGTTGTATGTAACCAAGGAAAGCGCAATCC 1546
QY 1787 AGGATG--GTAACAAACAGCGGCTGTTTGGATTGGAGCAGGCGATGTCATCCCGC 1843
DB 1547 TGGATGAAGATACAATATTGACGGGATTTCCGAGTTGGTGTGACATGTTAGTCCCT 1606
QY 1844 AAAAGGGGATAAAACCCGGGATTTGTTTACAACATTTCAACAGTGGATTAACATTAAC 1903
DB 1607 CTGAAGCCATTGACCCCGTCTGATCTATGACATTTGATGATGACAAATCATTTCAATC 1666
QY 1904 TCTGCATCTTTGGATTCAAGATCATATTTTAGCAATCACTCATPAAGACGTGAGCT 1963
DB 1667 TCTGTGCGCTGGGTTACACAGATTTTCAAGTTGAGATAATTGCAAAATCAGAAAGATGCCT 1726
QY 1964 GCATGGAATATTGCGGAAACCCGGGTTTGTCTCAATTAACCGTGCATAGCGGTGA 2023
DB 1727 GCAAGGAA-----GCAAAATAACTGAAGTGAACCTTAACATCACTAGTAGCAGTAA 1780
QY 2024 TTTTCAAGTGGCAAGACTACGAGATGATCAAGGCGGTGTCACTAAAGTTGGGAGTC 2083
DB 1781 GGGCAAGTGTGGCAAGCTTGTGTGAACCGGAC-----TGTTACTATGTGGGAGAAG 1834
QY 2084 CTAACTCGATATCTCAGTGAATGTAAGGCTCCAGAGGGGATCAAAGTTTATTTCAATC 2143
DB 1835 CAAATTCAGCTACACCGTGGAAATTTGATGCTCCTAGGAGGTGATGACAAAGTGTGTCA 1894
QY 2144 CTAAGAGACTTGTGTTCAACACAGTGTGATCAGAGCGTGACATATAGATGTTGTTAT 2203
DB 1895 CAACAAAGCTAGAAATTCACCAAGATGAAGGAGAGAAACATTTTCAATTAAGCTTAAGTT 1954
QY 2204 TGAAGAGAGAAACA 2218
DB 1955 GGGATATTAGCAAAA 1969

RESULT 6

US-10-260-238-485
; Sequence 485, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 485
; LENGTH: 2073
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-260-238-485

Query Match 9.1%; Score 211.6; DB 12; Length 2073;
Best Local Similarity 51.4%; Pred. No. 1.2e-55;

Matches 686; Conservative 0; Mismatches 609; Indels 39; Gaps 7;
QY 622 AACATGCTCTGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACATCACCGCC 681
DB 367 AATGAGACACAAGAGTCCAAAGTCACTCTCAGCACCGAAGCCACGGCACACACATGCA 426
QY 682 TCAACAGTTGGTGAATCTCTGTTTCGATGCGGAATGTTCTTGGCAATGAGCTGGTGTG 741
DB 427 TCTACCGCGGTGGTTACGCGGTCCCTGGCGCTAACTTTTGGTATGCGCAATGGCAC 486
QY 742 GCTCTGGGATGGGTCCCTGGAGCTCACATTTGCACTTATAAAGTCTGTGTGTTCAATGGT 801
DB 487 GCCCAAGCATGGCTATACGTGCACATTCATCTCAAAAGTATGCTGAGAGATGAT 546
QY 802 -----TGTTACAGCTCTGACATTTAGAGCTATAGATGATGAGCAATTCAGAT 849
DB 547 GGAATATCATCGTGGCTTACCTCTGACATTTCTCGCGGAATGAATGAGCAATTCAGAT 606
QY 850 AAGTTCGATGTTCTTCCCTTTCCCTGGCGTTTCCCTATTTCTTGTATGATGACACA 909
DB 607 GAGTAGATGTCAATATCCCTTTCCCTGGGTGTCTAAAGCCACAACTCTCAACGAAACC 666
QY 910 ATCGCATTTGGAACATTTCCGAGCCATGGAACGCGGTATATCTGTAATCTGTGACAGTGGT 969
DB 667 ACCTCCCTTGGCGAATTCATGCCATCCGTAGAGCATAGTCGTGTCCACCTCGCTGGA 726
QY 970 AACACCGTCCAAATCGAAAGCTCTGTTGCAAAACACAGCTCCTTGGGTCTCAACATTTGGC 1029
DB 727 AATGATGCCCTTGGCAGCTATCTACTTAATAACCTCGCACCATGGTGATAACGGTTGGC 786
QY 1030 GCAGGACGCTTGTATGCAAGATTTCCGCTGTGTGTCAGATTAGCCACGGGAAAGCTTCTC 1089
DB 787 GCATCAAGCATTTGACCGTCCGTTCCAGCTCACGTTGTCTTGGACATAATTCGGGAACC 846
QY 1090 TATGAGAGATCATTTGATCCGGGAAAGGTATGAAGATCCCGGAGAGAGGTTGAGGTG 1149
DB 847 TACATCGGCACCTCG---CTGATTTTCGCGCAAAACACAGCGGTTCATTTCTACCTCTA 903
QY 1150 ATTTACGTCAACAGGAGAGATAAAGGAGTGCAGTTCTGTTTGAGAGGTCACATTTCCAAGA 1209
DB 904 GTGTAC-----GGTGTGATGCTGGTCTGCTGTGTGATACGGGATGCTTTCCAGC 957
QY 1210 GAAGAAATCCAGGCAAAATGTTGATTTGTGATCGGAGTCAATGAA-----GATCG 1263
DB 958 AACATGTCACCGGAAAGATTTGCTCTGTGATGGCACAAAAATACTACCAACCCCAT 1017
QY 1264 GAGAAAGAGAGCGGTTAAGAGAGCTGGAGAGTTGCAATGATCTTAGCCATACAGAG 1323
DB 1018 GTGCAAGAGCTGCTGTCTCAACAAAGCGGTTGGTGTGGTGCATCATTTCAATTTGCTCCA 1077
QY 1324 ATCAACCAAGAGAGATTTCTATTGAGCTTCACTCTTACCAGCTACATTGATTTGTTAC 1383
DB 1078 GAATATGTTGATTTCTTACAGAGCTTCGCTGATATCTTGCCAACTGACGATTTACCTTC 1137
QY 1384 ACTGAGTCAGTCTTCTGAAGGCTTATGTTAATCCACGGTGAACCAACGAGCGCGGATA 1443
DB 1138 AAGGACACCGAAACGATCCATTCGTACACAGTCGGTGGCGCCGCGTTCGAAGAATC 1197
QY 1444 ATTTTGTGTTGATGATTTGAGGAGTGCAC---GAGCACGGAGGTGGTCAAGTTTCA 1500
DB 1198 GACTTCTTAGGCACTGTGATCAACCAATCACCGTCTGTCTCCAAAGGTCGACGCTTCTCA 1257
QY 1501 GCTCGAGGACCGGTTTAGCCAACTCTCGATATAAACCAGGATATGATTTGCTCCGGA 1560
DB 1258 AGTCGTGGACCGAACCCGCTTTGCTCCAGAGATCTCAAGCCCGCATGATTTGCCCGCGGT 1317
QY 1561 GTCATATCATTTGGGCTTGGCTCAAAATCTAGGACCAACC---GGACTTCTCTTATGAT 1617
DB 1318 GTGCACTCTCGCTGCTGATGAGACCGGAGAAATGTGCTTACATGGCCCAATGTATCGAC 1377
QY 1618 TCAAGAGAGTTAATCTTCACTGTATGTAGGAACTTCAATGTCTTGTCCCACTGTAGC 1677
DB 1378 AATAGCGCTCGAGTTCAACATCATCTCCGCGCAGTCCATGGCGTGCCTCCACATGAGC 1437

QY 1678 GGAATCACTGCTTTATCCGGTCTGTCATACCCGAACTGGTCTCCAGCTGCAATCAATCC 1737
Db 1438 GGCATCGCCCTGATGCTTAAGTGGCAACACCGAGTTGAGCGCCAGCGCAATCAAGTCG 1497
QY 1738 GCATTGATGACAAACAGCGGATTTGTACGATCGTCAAGGAAAGCGATAAAGGATGGTAAC 1797
Db 1498 GCCATGATGAGGAGCGGCTTACATGATGGCAATGATGGCAATGCCATCAAGGACATGGCC 1557
QY 1798 A-----AACCGCCGGTGTGGTGGATGGAGCAGGCGCATGTGAATCCGCAAAAGCG 1851
Db 1558 ACGGGCAGGACGACCGAGCGGCTTGGAGCTTGGGCTCGGCCACGTCGACCCCAACCGTCC 1617
QY 1852 ATAAACCCGGATGGTTTACCAATTCACACAGTGGATTACATAACTTACCTCTGCACT 1911
Db 1618 CTGACCCCTGGATTGGTGAACACACACCGCGGACGATTATACATCTCTCTGCAC 1677
QY 1912 CTTGGATTCAACAG 1925
Db 1678 CTGCGCTACACAG 1691

RESULT 7

US-09-770-149-59

; Sequence 59, Application US/09770149

; Patent No. US20020059663A1

; GENERAL INFORMATION:

; APPLICANT: Gorlach, Jorn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Raines, Tracy M.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Matthew, Abraham V.

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Woessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Garcia, Carlos A.

; APPLICANT: Kriker, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Hurban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; TITLE OF INVENTION: thaliana

; FILE REFERENCE: 2024 (PARA-013PRV)

; CURRENT APPLICATION NUMBER: US/09/770,149

; CURRENT FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: 60/178,506

; PRIOR FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 999

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 59

; LENGTH: 740

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-770-149-59

Query Match 7.5%; Score 174.6; DB 9; Length 740;

Best Local Similarity 57.0%; Pred. No. 3.5e-44;

Matches 338; Conservative 0; Mismatches 252; Indels 3; Gaps 1;

QY 1165 GGATTAAGAAAGTGAATGTTCTGTTGAGAGGTCACCTCCAGAGAGAAATCCGAGCC 1224

Db 131 GCGGATTGCGGAGACAGATTGTGTACCTGGGAATTTGAATTCATCATTTGTTGAAGCC 190

QY 1225 AAAATGTTGATTTGTGATCGCGGAGTCAATGGAAGATCGGAGAAAGAGAGCGGTTAA 1284

Db 191 AAAATGTTGATTTGTGATCGGAGGAGCAACGCAAGAGTTGAGAAAGAGTGCATCAAG 250

QY 1285 GAAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCAAGAGAGATTCT 1344
Db 251 CTAGCCGGTGGTCTGGTATGATTTCTGGGAAACACAGCTGAAAGCGGTGAAGAAATTAACC 310
QY 1345 ATTGACGTTTCATCTCTTACCAGCTACATTGATTGGTTTACACTGAGTCAGTCTTCTGAAG 1404
Db 311 GCGGATTTCGATCTCGTCCCGGCAATGTTGGAGCTTAAGCTGGAGATCAATCCGC 370
QY 1405 GCTTATTTAATGCCAGGTGAAACCAAGGCGCGGATAATTTTGGTGGTACGGTGAAT 1464
Db 371 GACTACATCAAAACATCAGACTCTCCACTGCAAAAATCAGTTTCTTAGSCACTTTGATC 430
QY 1465 GGGAGGTACGA---GCACCGGAGGTGGCTCAGTTTTTTCAGCTCGAGGACCGAGTTTACGC 1521
Db 431 GGACCACTCTCTCTCTCCAGAGTCGCGGCTTTCTCCAGCGCTGGACCGAATCACTTG 490
QY 1522 AATCTTCGATACATAAACCGGATATGATGCTCCGGAGTCAATATCATTTGGCGCTTG 1581
Db 491 ACACCGGTTATTTCTTAAACCGGAGCTGATTGCTCCTGGAGTCAACATTTTAGCCGTTGG 550
QY 1582 CCTCAAAATCTAGGACCAACCGGACTTCTCTTATGATTCAAGAAAGAGTTAACTTCACTGTA 1641
Db 551 ACGGATGGTGGTCTCTACCGATTAGATATCGATCCAGAGCGGTTCAATTCACATC 610
QY 1642 ATGTCAAGAACTTCAATGCTTTGTCCACATGTTTAGCGGAATCACTGCTCTTATCCGGTCT 1701
Db 611 ATCTCCGTTACATCGATGCTGTCGCCACACGTTAGTGAATCGCGCTCTCTCCGTA 670
QY 1702 GCATACCCGAACTGCTCTCCAGCTGCATCAAAATCCGATTCGATGACACAGC 1754
Db 671 GCTCATCCGATTGGTCACTCGAGCAATCAAAATCCGCTTGTAAACCCCGC 723

RESULT 8

US-10-260-238-5684/c

; Sequence 5684, Application US/10260238

; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.

; APPLICANT: Moughamer, Todd G.

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Provart, Nicholas

; APPLICANT: Ricke, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION

; FILE REFERENCE: 6011-NP

; CURRENT APPLICATION NUMBER: US/10/260,238

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04

; NUMBER OF SEQ ID NOS: 6077

; SEQ ID NO 5684

; LENGTH: 606

; TYPE: DNA

; ORGANISM: Zea mays

US-10-260-238-5684

Query Match 6.6%; Score 153.2; DB 12; Length 606;

Best Local Similarity 55.4%; Pred. No. 2e-37;

Matches 296; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

QY 1221 AGGCAAAATGTTGATTTGATTCGCGAGTCAATGGAAGATCGGAGAAAGAGAGCGGT 1280

Db 537 AGGCAAAATGTTGATTCGCGAGTCAATGGAAGATCGGAGAAAGAGAGCGGT 478

QY 1281 TAAAGAGCTGGAGGAGTTCATATGATCTTAGCCAAATACAGATCAACAGAGAGA 1340
DB 477 CAAGGAAGCTGGTGGCGCGCATGTGTCTTACTAAACGGGAGGCCAACGGCGAGGACT 418
QY 1341 TTCTATTGAGCTTCATCTTTACAGCTACATTTGATTTGTTTACACTGAGTCACTTCTTCT 1400
DB 417 GATTGGAGAGCGCCACGTCCTCTCTGTTCCGCGGTGGGCGAGAGAGGCGACCGGT 358
QY 1401 GAAGGCTTATGTTTAAATCCACCGGTGAACCAAGAGGCGCGGATATTTTGGTGGTACGTT 1460
DB 357 CAAGCGTATGCGCGGAGCGCTCCAGCCGATGGGAGACAATCAGCTTGGGGGCGACGCT 238
QY 1461 GATTGGAGGTCAGAGACCGGAGGTGGCTCATGTTTTCAGTCTGAGGACCGAGTTTACG 1520
DB 297 CGTGGCGGTCAAGCGCGCACCGGTGGTGGCTCTCTCTCGCGCGGTGGAGCCCAACCGGCT 238
QY 1521 CAATCTTCGATATAAAGCGGATATGTTGCTCCGGGAGTCAATATCATTTCCGGCTTG 1580
DB 237 CGTCCCGGAGATTTCAAGCGCGGACTTCATCGCACCGGGGCTCAACATCTCGCCCGTG 178
QY 1581 GCCTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAGAGAGAGTTAACTTCACTGT 1640
DB 177 GACGGGCGCACCGGTCCGACCGGCTGGAGGGGATACCCGCGGACGAGTTCAACAT 118
QY 1641 AATCTCAGGAATTTCAATGTCTTGTCCACATGTTAGCGGAATCACTGCTTTATCCGGTC 1700
DB 117 CTTGTCGGAGAGTCCATGCGTCCCGGACGCGCGGCGCGCGCTGCTCCGGTC 58
QY 1701 TGATACCGGACTGGTCTCAGCTGCAATCAATTCGGATTCATGACACAGC 1754
DB 57 CGCGACCCCGGGTGGTCCGCGGGGATACGCTCGCGCTGATGACTACGGC 4

RESULT 9

US-10-260-238-703
; Sequence 703, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 703
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (354)..(354)
; OTHER INFORMATION: n = any nucleotide

Query Match 6.1%; Score 141.4; DB 12; Length 1467;
Best Local Similarity 51.7%; Pred. No. 2.1e-33;

Matches 378; Conservative 0; Mismatches 341; Indels 12; Gaps 2;
QY 395 GGTGATGCTTAATCTCGGTTTGGTCAAGGCAATTTATCGCGTCTGTGATCTGGA 444
DB 367 GGCCTGTGGGAGACGGCCAGCTACGGCGACGGCGTCTATCGTCCGCGTGGTGGACACCGGG 426
QY 445 GTTTGGCCTGAAAGTTCCTAGCTTTGACGATACCGGAATGCTTCTGATTTCCACGGAATGG 504
DB 427 GTTTGGCCGAGAGCGGAGCTACGGCAACGACGGGTCCGCGCGTCCGCGCGGTGG 486
QY 505 AAAGGGATTTCCCAAGAGAGAAAGTTTCA---GTTCTTCGAGCTGTAAACCGGAAGCTA 561
DB 487 AAGGGCTACTGGAGTCCGGCACCGGTTTCGACGGCGCCAAAGGGGTGCAACCGGAAGCTC 546
QY 562 ATCGGTGTGATTTCTTCAATCAGAGGACACCGTGTCTGATTAATTCACAGAGGAATCACCA 621
DB 547 ATCGGGGCGCAAGTTTCAGCGCGGCTCCGCCCGCGCT-----GGGAGGCGG 597
QY 622 AACATGCTCTGATATCATTTCCGCAAGAGATTCAACGGGACACGGGATTCACACCGCC 681
DB 598 AACATCAGATCGCGGTGAATTCGCCGCGGACACGGACGGCCACGGCACGACACCTCG 657
QY 682 TCAACAGTTGTTGATCTCTGTTCGATGGCGAATGTTCTTGGCAATGAGAGCTGGTGTG 741
DB 658 TCGACGGCGCGGCTCGCCGCTCCGGGGCGTCTTCTTCGCTACGCTCCCGGGGTG 717
QY 742 GCTCTGGGATGGTCTCTGGAGCTCACATTGCACTTATAAAGTCTGTGTTGTTCAATGGT 801
DB 718 GCACGGGCGATGGCGCCACGCGCGAGGTGGCGCTGTACAAAGTCTCTGTTTCGACGAGGGC 777
QY 802 TGTTACGCTCTGATCTTAGAGCTATAGATGTAGGATTCAGATAAAGTCTGATGTT 861
DB 778 GGGTACAGCGGACATCGTCGCGCCATAGACCAAGCCATCGCAGACGGGCTCGACGTG 837
QY 862 CTTTCGCTTTCCTTGGCGGTTTCCTATTCTTGTATGATGACACAATCGCCATGGA 921
DB 838 CTCTCCATCTCGTGGGCTTCAACACCGCCGCTCCACCGACCTGTGTCATCGGC 897
QY 922 ACATTCGAGGATCGGAACGGGTATATCTGTAATCTGTGAGCTGTGTAAACAGGTCCA 981
DB 898 TCGTTGCGCGCAATGACGACGGCATCTTCGTGTCACCTCAGCGGCAACGACGGCCG 957
QY 982 ATCGAAAGCTCTGTTGMAAACACAGCTCTTGGGTCTCAACATTGGCGGAGGACCGTT 1041
DB 958 GGCCTCTCCGCTCTCCACACGGCGCGCTGGGCACTGACCGTCCGCGCTGGCACCGTG 1017
QY 1042 GATCGAAGTTTCCGCTGTGTCAGATTAGCCCAACGGAAGCTTCTCTATGTGAGAGTCA 1101
DB 1018 GACAGGGAGTTCTCCGGCATCTCGAGCTCGGATGCGACCGCTATCGGGGAATCG 1077
QY 1102 TTGTATCCGGG 1112
DB 1078 CTGTACGCGG 1088

RESULT 10

US-10-260-238-1133
; Sequence 1133, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP

FILE REFERENCE: 2026 (PARA-015PRV)
CURRENT APPLICATION NUMBER: US/09/770,961
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,456
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 228
LENGTH: 510
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-961-228

Query Match 4.4%; Score 103.2; DB 11; Length 510;
Best Local Similarity 54.5%; Pred. No. 1.5e-21;
Matches 207; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 200 CTTCTTCTCGACTTCTCTACTCCTTGTGCTCTGCTGATGAAGATTGCTGCTCAGTTGA 259
DB 128 CTCCTCCCTCCATTATCCACATACGACACCGTTTCCATGGCTTCTCCGCTAGACTCA 187
QY 260 CTGAATCAGAACCGGAGATCTGAGATATTCACCTGAAGTTGTTGCACTGAGACCTGACC 319
DB 188 CTTTCAAGAGCGTAGTCACTCTTACACCATCTCTCATGTCATCTCTGTTATCCGGAAC 247
QY 320 ATGTTCTTCAAGTTCAAAACCACTTACTCTTACAAAGTTCTTGGGACTCGACGGTTTGGAA 379
DB 248 AAGTCCGTCACCTTGCACACCACTCGTCCCTCGAGTTCTTGGTCTTAGTCCACCGACA 307
QY 380 ACTCGGTGTATGTCTAAATCTCGGTTTGTGTAAGGACAAATATCGGGTGTCTGTGATA 439
DB 308 AAGCGGTCTACTCGAAGAGTCTGATTTCCGGTCCGACTAGTATATCGGAGTTATCGATA 367
QY 440 CTGAGTTTGGCTTGAAGTCTTACGTTTTCAGCATCCGGAATGCCCTTCGATCCACGGA 499
DB 368 CTGTTGTTGGCCGGAAGACCTAGCTTTGATGACCGTGTCTTGGTCTTCTTCCCATTA 427
QY 500 AATGGAAGGATTTGGCAAGAGGAGAAAGTTTTCAGTTCTTTCAGCTGTAAACCGGAAGC 559
DB 428 AATGGAAGGCAATGATCGTCCCAAGATTTCCGGAGTCTGCTTGAACGGTAAC 487
QY 560 TAATCGGTGTAGATTTCTTC 579
DB 488 TCGTCCGAGCTAGATTTCTTC 507

RESULT 12
US-09-938-842A-338
Sequence 338, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 338
LENGTH: 2958
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-338

Query Match 4.3%; Score 100.2; DB 10; Length 2958;
Best Local Similarity 56.5%; Pred. No. 4.7e-20;
Matches 186; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 1425 GAAACCAAGCGCGGATAATTTTGGTGGTACGGTGAATGGGAGGTCCAGGACACCGGA 1484
DB 1701 GAAGCGGATTGCACACATTACTCTTCAAGGACAGATTGGGACTAAACACGCGCTGT 1760
QY 1485 GGTGCTCAGTTTTCAGTCCGAGGACCGAGTTTAGCCAAATCTTCGATACATAAAACCGGA 1544
DB 1761 AATGGCTTCATTTTCTTCAAAAGGTCACAGCATAGTGGCTCTCAGATTCTGAAGCTGA 1820
QY 1545 TATGATTGCTCCGGAGTCAATATCATTTGGCGCTGGCTCAAAATCTAGACCAACCGG 1604
DB 1821 CATAACTGCTCTGCTGTGAGTGTATCGCTGCTTACACTGGAGCAGTCTCTCCAACAAA 1880
QY 1605 ACTTCTTATGATTTCAAGAAGAGTTAACTTCACTGTAATGTGAGAACTTCAATGCTTTG 1664
DB 1881 CGAACAGTTTGAATCTCGACGACTTCTGTTCAATGCTAATTCAGGAACCTTATGCTTTG 1940
QY 1665 TCCACATGTTAGCGGAATCACTGCTTCTTATCCGGTCTGCATACCCGAACTGGTCTCCAGC 1724
DB 1941 TCTTCATATCTCGCAATTCGGGCTTCTCAAAACTCGTTATCTTCTTGGAGCCCGC 2000
QY 1725 TCGAATCAAAATCCGCAATTGATGACAAAG 1753
DB 2001 AGCTATCGGCTCTGCCATCATGACTACCG 2029

RESULT 13
US-09-938-842A-338
Sequence 338, Application US/09938842A
Patent No. US20040009476A9
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 338
LENGTH: 2958
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-338

Query Match 4.3%; Score 100.2; DB 12; Length 2958;
Best Local Similarity 56.5%; Pred. No. 4.7e-20;
Matches 186; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 1425 GAAACCAAGCGCGGATAATTTTGGTGGTACGGTGAATGGGAGGTCCAGGACACCGGA 1484
DB 1701 GAAGCGGATTGCACACATTACTCTTCAAGGACAGATTGGGACTAAACACGCGCTGT 1760
QY 1485 GGTGCTCAGTTTTCAGTCCGAGGACCGAGTTTAGCCAAATCTTCGATACATAAAACCGGA 1544
DB 1761 AATGGCTTCATTTTCTTCAAAAGGTCACAGCATAGTGGCTCTCAGATTCTGAAGCTGA 1820
QY 1545 TATGATTGCTCCGGAGTCAATATCATTTGGCGCTGGCTCAAAATCTAGACCAACCGG 1604
DB 1821 CATAACTGCTCTGCTGTGAGTGTATCGCTGCTTACACTGGAGCAGTCTCTCCAACAAA 1880
QY 1605 ACTTCTTATGATTTCAAGAAGAGTTAACTTCACTGTAATGTGAGAACTTCAATGCTTTG 1664

Db 1881 CGAACAGTTGATCTCTCGACGACTTCTGTTCAATGTTATTTACGAACTCTATGCTTG 1940
Qy 1665 TCACATGTTAGCGGAATCACTGCTTTATCCGGTCTGCATACCCGAACTGGTCTCCAGC 1724
Db 1941 TCCTCATATCTTGGCAATTCGGGGTCTTCTCAAACTCGTTATCCTTCTTGGAGCCCGC 2000
Qy 1725 TGCATCAAAATCCGCAATGATGACACAG 1753
Db 2001 AGCTATCCGCTCTGCCATCATGACTACCG 2029

RESULT 14
US-10-260-238-3023
; Sequence 3023, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 3023
; LENGTH: 849
; TYPE: DNA
; ORGANISM: *Triticum aestivum*
; FEATURE:
; NAME/KEY: N region
; LOCATION: (814)..(814)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-3023

Query Match 4.2%; Score 98.4; DB 12; Length 849;
Best Local Similarity 51.0%; Pred. No. 7.3e-20;
Matches 358; Conservative 0; Mismatches 311; Indels 33; Gaps 4;
Qy 1496 TTTGAGTTCGAGGACCGAGTTAGCCAACTCTTCGATCACTAAACCGGATGATGCTC 1555
Db 1 TCTTTCCGGTGGTCCAGCGTCTCGCCCTCGCATCTTAAGCCGGATATATTGGCAC 60
Qy 1556 CGGAGTCAATATCATTTGGCGCTTGGCTTCAAAATCTAGACCAACCGGACTCTCTTATG 1615
Db 61 CAGGCCTCAATATCTCGCCCGTGGCCGCGCCAGCAGC----- 101
Qy 1616 ATTCAGAGAGTTAACTTCACTCTAATGTCAGAACTTCAATGCTTGTCCACATGTTA 1675
Db 102 --TCTGGAAGGGGCGCCCTTTCAACATCATATCAGGAACATCATGGGACGCCACAGTCA 159
Qy 1676 GCGGAATCACTGCTTTATCCGGTCTGCAATCCCGAACTGGTCTCCAGCTGCAATCAAAAT 1735
Db 160 GTGGTGGCTTGCATTTATCCAAAGCATCTCTGACTGGTGGCCAGCGCCATCAAGT 219
Qy 1736 CCGCATTTGATGACACACGCGGATTTGTACGATCTGTCAAGGAAAGCGGATTAAGATG--- 1792
Db 220 CTGCATCTTAACGACGTCGGACATCGTTAATAGCACTGGTGGCTCAATCTTGGACGAGC 279

Qy 1793 GTAACACACGCGGTGTTTTCGATTGAGCAGGCGCATGTGAATCCGCAAAAGCGCA 1852
Db 280 AACATAGGAAGCGCGCGTGTACACACAGCGCGCGCCATGTGAACCCGACGAGCGC 339
Qy 1853 TAAACCCGGGATTCGTTTACAAATTCACACAGTGGATTACATAACTTACCTCTGCACTC 1912
Db 340 CAGATCTTGGCTTGGTGTACGACCTCAGTGTCTAGTACTAGCGCGGCTACATCTGTGGC 399
Qy 1913 TTGGATTCACAAGATCAGATATTTTAGCAATCACTATAAGAACGTCGAGCTGCAANTGAA 1972
Db 400 TCCTAGGTGACAG--CGGCCTAGCAACCTCGTGCACAACTCAAGCTTGACCTGTGCAA 456
Qy 1973 TATTGGGAAAAACCCCGGTTTGTCTCAATTAACCGTGCATAGCCGCTGATTTTCAAC 2032
Db 457 AGCTGCCCAAGGTCCACGCGCTTAACTACCAGCATAAACCGTGTACGCGACA- 515
Qy 2033 GTGGCAAGACTACGAGATGATCAAGGCGTGTCACTAACGTTGGGAGTCTTAACCTCGA 2092
Db 516 -----TCGACGCCATTACAGTAAGCGGACTGTGACAAACGTTGGCCAGCAACTCGA 570
Qy 2093 TATACTCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTTGTCAATCTTAAGAGAC 2152
Db 571 CATTACAGTGAAGTGTGACGCGCAAGATCACTGACAGTCCGTGTCTCCCGGAGACGC 630
Qy 2153 TTGTTTCAACACAGTGGATCAGAGCTGAGCTATAGATAT 2194
Db 631 TGGCATTCTCCAAAACCGGAGAGAAAGACCTTTCAGCGTGT 672

RESULT 15
US-10-260-238-3022
; Sequence 3022, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 3022
; LENGTH: 828
; TYPE: DNA
; ORGANISM: *Triticum aestivum*
US-10-260-238-3022

Query Match 4.0%; Score 92.4; DB 12; Length 828;
Best Local Similarity 51.2%; Pred. No. 5.9e-18;
Matches 297; Conservative 0; Mismatches 271; Indels 12; Gaps 3;
Qy 1618 TCAAGAGAGTTAACTTCACTGTATGTCAGGAATCTCAATGCTTGTCCACATGTTAGC 1677
Db 64 TCTGGAAGGGGCGACTTTCAACATCATATCAGGAACATCCATGGCAGCCACAGCTCAGT 123
Qy 1678 GGAATCACTGCTTTATCCGGTCTGTCATPACCCGAACTGGTCTCCAGCTGCAATCAAAATCC 1737
Db 124 GGTGTCGCTGGCTTATCAAAAGCATCCATCTGACTGTCGCGCAGCGCCATCAAAATCT 183


```
QY 1738 GCATTGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAAGCGATAAAGGATG---GT 1794
|||
Db 184 GCCATCCTAACGACGTCGGACATCGTTAATAGCACTGGTGGCTCAATCTTGGACGAGCAA 243
|||
QY 1795 AACAAACCCAGCCGGTGTGTTGGATGGAGCAGGCGATGTGAATCCGCAAAAGGGATA 1854
|||
Db 244 CATAGGAAGCCGGCGTGTATGACACAGGCGCGCGCCCATGTGAACCCGACCGAGCCGCA 303
|||
QY 1855 AACCCGGGATTTGTACACATTCAACCACTGATTAACATTAACCTTACCTCTGCACTCTT 1914
|||
Db 304 GATCCTGGCTTGGTGTACGACCTCAGGCTCACCAGCTACCGGGCTACATCTGCTGGCTC 363
|||
QY 1915 GGATTCAACAAGATCAGATATTTTAGCAATCACTCATAGAAGCGTGAGCTGCAATGGAATA 1974
|||
Db 364 CTCGGTGACAGC---GGCCTAGCAACCATCGTGGCAACTCAAGCTTGACCTGGCGAAG 420
|||
QY 1975 TTGGGGAACCCGGGTTTTAGTCTCAATTACCCGTCGATAGCCGTGATTTTCAAACGT 2034
|||
Db 421 CTGCCCAAGGTCCATGCTGTGCAGCTTAACACCGACGATAACCGTGTACGCGACATC- 479
|||
QY 2035 GGCAAGACTACGGAGATGATCACAGGCGTGTCACTAAGCTTGGGAGTCTTAACCTCGATA 2094
|||
Db 480 -----GACGCCATTCAAGTGAAGCGGACTGTGACGAACGTTGGCCAGCAAACTCGCG 534
|||
QY 2095 TACTCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAGTTATTGTCAATCTTAAGAGACTT 2154
|||
Db 535 TTCACGGCGAAGGTGGAGCGGCCCAAGATCACTGACAGTGGGTCTCTCCCGGAGAGCGCTG 594
|||
QY 2155 GTGTTCAAACACGTGGATCAGACGCTGAGCTATAGAGTAT 2194
|||
Db 595 GGATTCTCCAAAACCGGAGAGAGAGACCTTCAGCGGTGT 634
|||
```

Search completed: February 3, 2004, 02:18:36
Job time : 630 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 21:36:20 ; Search time 3282 Seconds
(without alignments)
17239.738 Million cell updates/sec

Title: US-09-806-767-1
Perfect score: 2328
Sequence: 1 atggaacccaacacctttt.....taacctgaagactaactga 2328

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 23781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpi:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	574.8	24.7	832	28	BH480160
c 2	475	20.4	696	28	BH660501
c 3	472.4	20.3	676	28	BH998381
c 4	461	19.8	670	28	BH464077

C	5	438.2	18.8	635	28	BH599290
C	6	426.8	18.3	627	28	BH597572
C	7	424	18.2	640	28	BH442444
C	8	409.2	17.6	803	28	BH601308
C	9	356.2	15.3	527	28	BH714875
C	10	326.8	14.0	767	28	BH793726
C	11	296.8	12.7	568	10	BF052773
C	12	273	11.7	772	28	BH601319
C	13	270.2	11.6	789	29	CC340521
C	14	252	10.8	750	28	AZ046831
C	15	233.8	10.0	713	28	BH720365
C	16	228	9.8	412	9	AV420480
C	17	223.6	9.6	719	13	BQ870650
C	18	219.4	9.4	848	14	CD575777
C	19	217.6	9.3	761	28	BH442432
C	20	215.4	9.3	1572	11	AY107161
C	21	207.6	8.9	738	13	BQ996378
C	22	206.6	8.9	697	13	BU895782
C	23	205	8.8	790	28	BH590573
C	24	203.6	8.7	762	28	BH649984
C	25	201.8	8.7	863	10	BF278122
C	26	199	8.5	735	13	BQ893301
C	27	198	8.5	800	14	CB892659
C	28	197.4	8.5	864	28	BH648619
C	29	196.8	8.5	766	28	BH725669
C	30	195.6	8.4	703	13	BQ999661
C	31	193.8	8.3	719	28	BH668947
C	32	190.2	8.2	662	28	BH434517
C	33	190.2	8.2	734	13	BQ866897
C	34	189	8.1	926	10	BG445439
C	35	188.8	8.1	610	13	BH883192
C	36	188.8	8.1	639	9	AW720622
C	37	188.4	8.1	698	28	BH994888
C	38	187.8	8.1	716	28	BH659528
C	39	186.8	8.0	526	9	AV527216
C	40	186.4	8.0	640	10	BF005920
C	41	184.8	7.9	740	13	BQ867168
C	42	184.4	7.9	558	28	BH714885
C	43	183	7.9	709	13	BQ866922
C	44	183	7.9	818	12	BI405859
C	45	182.8	7.9	639	9	A1730592

ALIGNMENTS

RESULT 1
BH480160/c

LOCUS

DEFINITION

BOHME48TF BOHM Brassica oleracea genomic clone BOHME48, genomic

survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH480160 832 bp DNA linear GSS 13-DEC-2001
BOHME48TF BOHM Brassica oleracea genomic clone BOHME48, genomic
survey sequence.

BH480160 GI:17698264

GSS.

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Brassica

1 (bases 1 to 832)

Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Other_GSSs: BOHME48TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

```

FEATURES
  source
    Location/Qualifiers
      1..832
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="TO100DH3"
        /db_xref="taxon:3712"
        /clone="BOHME48"
        /clone_lib="BOHM"
        /notes="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
        genomic DNA inserted into PHOS1 using BstXI linkers"
      BASE COUNT 177 a 209 c 247 g 199 t
      ORIGIN
        Query Match 24.7%; Score 574.8; DB 28; Length 832;
        Best Local Similarity 81.9%; Pred. No. 7.2e-129;
        Matches 663; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
      QY 248 CTGCTGAGTTGACTGAATCAGAAGCCGAGATACAGATATTCACCTGAAGTTGTTGCG 307
      Db 831 CAGCTCAGTTAAACGAATCAGAAGCCAAACACTCAAGACTTTACCTGAAGTTGTTGCG 772
      QY 308 TGAGACCTGACCATGTTCTTCAGGTTCAACACCACTTACTCTTACAGATTCTTGGGACTCG 367
      Db 771 TAAGACCTGACCATGTTCTCAAGTCCAAACCACTTACTCTTACAGATTCTTGGACTCA 712
      QY 368 ACGGTTTGGAAACTCCGGTGTATGTTCTAAATCTCGGTTTGGTCAAGGCACAAATTATCG 427
      Db 711 ACGGTCAGACCCCTCCAGCGTATGTTCTAAATCCCGGTCGGCCAGGCACGATCATCG 552
      QY 428 GCGTGTGATATCGAGTTTGGCTGAAAGTCTTAGCTTTGACGATACCGAAATGCTTT 487
      Db 651 GCGTTCGACACAGAGTCTGCGCCGAAAGCCCTAGCTTCGACGACACAGGATGCTTT 592
      QY 488 CGATTCCACGGAATGGAAGGATTTGCCAAGAGGAAAGTTTCAGTTCTTCGAGCT 547
      Db 591 CAGTCCCAAGCAATGGAAGGTTTTCGCAAGAGGAGAGCTTCCTTCTCAACT 532
      QY 548 GTAACCGGAAGTAAATCGGTGTAGATTCTTATCAGAGCACACCGTGTGCTTAATTAC 607
      Db 531 GTAACAGAAGAACTAATCGGCGCTAGATTCTTATCAGAGCCACCGCTAGCCACTTC 472
      QY 608 CAGAGGAATCACCACCAATCCCTCGTGAATACATATTTCCGCAAGAGATTCAACGGGACAG 667
      Db 471 CTTTGGACTCACCACCAATCCCTCGGATACATATCGGCAAGAGACTCCACGGGSCAG 412
      QY 668 GGAATCACAGGCTCAACAGTTGTTGGATCCTCTGTTTTCGATGGGGAATGTTCTTGGCA 727
      Db 411 GGACCCACACCGGCTCAACTGCTGGCGGCTCTCTGTTTCGACGGCGAGCGTTCTCGTA 352
      QY 728 ATGAGCTGTTGGCTCGTGGATGCTCTCTGAGCTCACATTGCACTTAAGTCT 787
      Db 351 ACGAGCCGGGTAGCCCGGGATGGCCGAGAGCTACGTGGCGGTTTCAAAAGTCT 292
      QY 788 GTTGGTTCAATGTTGTTACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCGATTCAAG 847
      Db 291 GCTGTTCAACGCTGTTACAGCTCAGACATTTGCGGGGATAGAGTAGCGATACAAG 232
      QY 848 ATAAAGTCGATGTTCTTTCGCTTCCCTTGGCGGTTTCCCTATTCCTTTGTATGATGACA 907
      Db 231 ACAAGTCGACGTTCTATCCCTCTCCCTAGCGGTTTCCCTATCCCTTTGTATGATGACA 172
      QY 908 CAATCGCCATTGGACATTCGAGCCATGGAACCGGTTATATCTGTAATCTGTGAGCTG 967
      Db 171 CTATCGCCATTGGTACCTTCGCGCCACGGAATGGAATCTCCGTGTGTCGCGGGTG 112
      QY 968 GTAACACGGTCCCAATCGAAAGTCTGTGCAACACACAGCTCTCTGGGTCTCAACATTG 1027
      Db 111 GCAACACGGCCGCTCGCTAGCTCGGTGGGGAACACAGCTCTCTGGGTCTCAACATTG 52
      QY 1028 GCGCAGGACGCTTTGATCGAAGATTTCCCG 1057
      Db 51 GAGCAGGAACGCTCGATAGAAAATTCGCG 22
  
```

```

RESULT 2
BH660501
LOCUS BOHWS71TF BO_2_3 KB Brassica oleracea genomic clone BOHWS71,
        696 bp DNA linear GSS 19-FEB-2002
        genomic survey sequence.
ACCESSION BH660501
VERSION BH660501.1 GI:18719169
SOURCE GSS.
ORGANISM Brassica oleracea
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
          ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 696)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Chris Town
        TIGR
        9712 Medical Center Drive, Rockville, MD 20850, USA.
        Tel: 301-838-3523
        Fax: 301-838-0208
        Email: cdtown@tigr.org
        DNA is from a doubled haploid provided by Tom Osborn.
        Seq primer: TF
        Class: sheared ends.
FEATURES
  source
    1..696
      /organism="Brassica oleracea"
      /mol_type="genomic DNA"
      /strain="TO100DH3"
      /db_xref="taxon:3712"
      /clone="BOHWS71"
      /clone_lib="BO_2_3 KB"
      /note="vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
      genomic DNA inserted into PHOS1 using BstXI linkers"
    BASE COUNT 163 a 166 c 220 g 147 t
    ORIGIN
      Query Match 20.4%; Score 475; DB 28; Length 696;
      Best Local Similarity 80.5%; Pred. No. 1.3e-104;
      Matches 556; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
    QY 685 ACAGTTGGTGGATCCTCTGTTTCGATGGCGAATCTTCTTCGCAATGAGCTGGTGTGCT 744
    Db 6 ACTGCTGGGCTCCTCTGTTTCGACGGCGAGGCTTCGCTAACGGAGCCGGGTAGCC 65
    QY 745 CGTGGGATGGCTCCTGGAGCTCACATTGCGATCTATAAAGTCTGTGTTCAATGGTTGT 804
    Db 66 CGCGGATGGCGCCAGAGCTCAGTGGCGGTTTACAAAGTCTGCTGTTCAACGGCTGT 125
    QY 805 TACAGCTCTGACATTCCTAGCAGCTATAGATGTAGCGATTCAAGATAAAGTCGATGTTCTT 864
    Db 126 TACAGCTCAGACATTCCTGGCGCGATAGAGTACGCGATACAAAGTCTGAGCTTCTA 185
    QY 865 TCGCTTTCCCTTGGCGGTTTCCCTATTCTTGTATGATGACACAAATCGCATTTGGACA 924
    Db 186 TCCCTCTCCCTAGCGGTTTCCCTATCCCTTGTATGATGACACTATCGCATTTGGTACC 245
    QY 925 TTCGAGCGCATGGAAACCGGTTATATCTGTAATCTGTGAGCTGTGTAACACGGTCCCAATC 984
    Db 246 TTCGCGCCACGGAACATGGAATCTCGTTGTCGCGGGTGGCAACACGGCCGCTC 305
    QY 985 GAAAGCTCTGTTCGAAACACAGCTCTTGGGTCTCAACCAATTTGSCGCGAGGACGCTTGAT 1044
    Db 306 GCTAGCTCGGTGGCGAACACAGCTCTTGGGTCTCAACCATCGGAGCAGGAACGCTCGAT 365
    QY 1045 CGAGATTTCCGCTGTGGTCAGATTAGCCAAACGGAAGCTTCTCTATGGAGATCAATG 1104
    Db 366 AGAAAATTTCCCGGGGTTTGGGTAGTCCAAACGGAAGCTTACTCTACGAGAGTCTGTTG 425
    QY 1105 TATCCGGAAAAGGTATAAAGAAATGCCGGGAGAGAGGTTGAGGTGATTACGTACACAGGA 1164
  
```

```

Db      426  TACCAGGAAGGATTAAGAGAGGCTGAGAGAGTTCGAGATTCTACGTGGGGGA 485
Qy      1165  GAGATTAAGAGAGTGAAGTCTTTGAGAGGCTCACTTCCAGAGAGAAATCCGAGGC 1224
Db      486  GGAGACAAAGAGAGAGTTCGTTGAGAGGCTCGCTCCGAGAGAGAGTGTCCTCAAGGC 545
Qy      1225  ABAATGGTGATTCTGATCCGAGGCAATGGAAGATCGAGAGAGAGAGAGCGGTAA 1284
Db      546  AAGATGGTGATATCGGACCTGGAGTCAACGGAGATCAGAGAGAGAGAGCGGTAA 605
Qy      1285  GAAGCTGGAGAGTTCGAATGATCTTAGCCAAATACAGAGATCAACCAAGAGAGATTCT 1344
Db      606  GAAGCTGGCGGCTGGAGATGATCTTGGCTAAACCGAGATTAAACCAAGAGAGATTCT 665
Qy      1345  ATTGACGTTTCATCTTACCAGCTACATTGA 1375
Db      666  GTAGACGTTTCATTGATACCAGCTACGTTGA 696

RESULT 3
BH998381/c
LOCUS   oeh58ell.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION
ACCESSION   BH998381
VERSION     BH998381.1 GI:23544689
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 676)
AUTHORS     Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE       Whole genome shotgun reads from Brassica oleracea
JOURNAL     Unpublished
COMMENT     Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Plate: oeh58 row: e column: 11
Seq primer: -21UPpOT forward
Class: shotgun
High quality sequence start: 14
High quality sequence stop: 551.
FEATURES    source
            1..676
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /db_xref="taxon:3712"
            /clone_lib="B.oleracea002"
            /notes="vector: pOTW13; Whole genome shotgun library from
            flowering buds. DNA was purified from a crude nuclear
            prep using Brassica oleracea TO1000DH3 buds provided by
            Thomas Osborn at the University of Wisconsin. Genomic
            DNA was provided by Pablo Rabinowicz (CSHL) and the
            shotgun library prepared at Washington University Genome
            Sequencing Center."
BASE COUNT  138 a 177 c 208 g 153 t
ORIGIN
Query Match 20.3%; Score 472.4; DB 28; Length 676;
Best Local Similarity 81.3%; Pred. No. 5,7e-104;
Matches 548; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
Qy      351  CAAGTCTCTGGAGCTCAGCGGTTTGAAACTCCGGTGATGCTCTAAATCTCGGTTGG 410
Db      676  CAAGTCTCTAGGACTCAACGGTCCAGGACCTCCAGCGTATGCTCTAAATCCCGGTC 617
Qy      411  TCAGGCACAATATCCGCGCTGTTGATCTGAGATTGGCCCTGAAAGTCTAGCTTGA 470

```

```

Db      616  CCAAGGCAGGATCATCGGCGTTCTCGACACAGGAGTCTGGCCCGAAAGCCCTAGCTTGA 557
Qy      471  CGATACCGGAATGCTTCGATTCCACGGAATGAAAGGATTTGCCAAGAGAGAGAAAG 530
Db      556  CGACACAGGGGATGCTTCAGTCCCAAGCAAATGGAAGGTGTTTTGCCAAGAGAGAGAG 497
Qy      531  TTTCAAGTCTTCGAGCTGTAACCGGAAGCTAATCGGTGCTAGATTCTTCATCAGAGACA 590
Db      496  CTTCACTTCTCAAACTGTAAACAAGAACTAATCGGCGCTAGATTCTTCATCAGAGGCCA 437
Qy      591  CCGTGTGCTGAATTCACCAGAGGAATCACCAGAAATGCTCGTGAATACATTTCCGCAAG 650
Db      436  CCGGCTAGCAACTCTCTTTGGACTCACCAGAAATGCTCGGCAATACATATCGGCACG 377
Qy      651  AGATTCAACGGGACACGGGACTCAACGGCTCAACAGTTGGTGGATCCTCTGTTTCGAT 710
Db      376  AGACTCCACGGGGCAGCGGACCCACACGGGCTCAACTGCTGGCGGCTCCTCTGTTTCGAC 317
Qy      711  GGCGAATGTTCTTGGCAATGGAGCTGGTGGCTCGTGGGATGGCTCTGGAGCTACAT 770
Db      316  GGCGAGCGTTCTCGGTAACGGAGCCGGGTAGCCCGGGGATGGCGCAGAGCTCAGCT 257
Qy      771  TGCAGTCTATAAAGTCTGTTGGTTCAATGGTTGTTACAGCTCTGACATTTCTAGCAGCTAT 830
Db      256  GCGGTTTACAAAGTCTGCTGTTCAACGGCTGTTACAGCTCAGACATTTCTGGCGGCGAT 197
Qy      831  AGATGTAGCGATTCAAGATAAAGTCGATGTTCTTTCGCTTTCCTTCGCGGTTTCCCTAT 890
Db      196  AGACGTAGCGATACAAGACAAAGTCGACGTTCTATCCCTCTCCCTAGCGGTTTCCCTAT 137
Qy      891  TCCTTTGTATGATACACATCGGCATGGAACATTCGAGGCCATGGAACCGGTATATC 950
Db      136  CCTTTGTATGATGACACTATCGCCATTGGTACCTTCCGCGCCACGGAATGGAATCTC 77
Qy      951  TGTATCTGTGCACCTGCTACACGCTGTCACACGCTGCCAATCGAAAGCTCTGTTCAACACAGCTCC 1010
Db      76  CGTTGCTGCGCGGGTGGCAACACGGCCGCTCGTAGCTCGGTGGCGAACACAGCTCC 17
Qy      1011  TTGGTCTTCAACCA 1024
Db      16  TTGGGATTCACCA 3

BH464077/c
LOCUS   BH464077.1 GI:17659406
DEFINITION   BH464077 BOGE Brassica oleracea genomic clone BOGEF01, genomic survey sequence.
ACCESSION   BH464077
VERSION     BH464077.1
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 670)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished
COMMENT     Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: Sheared ends.
FEATURES    Location/Qualifiers
            1..670
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"

```

```
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOGEF01"
/clone_lib="BOGE"
/notes="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 137 a 188 c 186 g 159 t
ORIGIN
Query Match 19.8%; Score 461; DB 28; Length 670;
Best Local Similarity 80.6%; Pred. No. 3.4e-101; Indels 0; Gaps 0;
Matches 539; Conservative 0; Mismatches 130;
Qy 472 GATACCGGATGCTTCGATTCACCGAAATGGAAGGATTTGCCAAGAGGAGAAAGT 531
Db 670 GACACAGGATGCTTCAGTCCCAAGCAATGGAAGGTTTGCCAAGAGGAGAGAGC 611
Qy 532 TTCAAGTTCCTCGAGCTGTAACCGGAAGCTAATCGGTGCTAGATTTCTTCATCAGAGACAC 591
Db 610 TTCACITTCCTCAAACTGTACAAAGAACTAATCGGCGCTAGATTTCTTCATCAGAGGCCAC 551
Qy 592 CGTGTGCTAATTCACAGAGGATCACCAGAACTGCTCGTGAATACATTTTCCGCAAGA 651
Db 550 CGCGTAGCCAACTCTCTTTGGACTCACCAGAACTGCTCGGGAATACATATCGGCACGA 491
Qy 652 GATTCAACGGGACACGGGACTCACACGGCTCAACAGTTGGTGGATCCTCTGTTTCGATG 711
Db 490 GACTCCACGGGGACACGGGACCCACACGGCTCACTGCTGGGGCTCTCTGTTTCGAGC 431
Qy 712 GCGAATGTTCTTGGAATGAGCTGTGTGGTCTGTGGAGTCTCTCGAGCTCACAT 771
Db 430 GCGAGCGTTCTCGTAAACGGGCGGGTAGCCGCGGATGGCGCAGGAGCTCACGTG 371
Qy 772 GCGAGCTATAAAGTCTGTTGGTCAATGGTTGTATAGCTCTGACATTTACAGCTATA 831
Db 370 GCGGTTTACAAAGTCTGCTGGTTCAACGGCTGTACAGCTCAGACATTTGCGGCGATA 311
Qy 832 GATGTACGATTCAGATAAAGTCGATGTTCTTTCCTTTCCTTGGCGGTTTCCCTATT 891
Db 310 GAGTAGCGGATACAGACAAAGTCGAGCTTCTATCCCTCTCCCTAGGCGGTTTCCCTATC 251
Qy 892 CCTTTGATGATGACACAAATCGCCATTGGAAATTCGAGCCATGGAACGCGGTATATCT 951
Db 250 CCTTTGATGATGACACTATCGCCATTGGTACCTTCCGCGCACGGAACATGGAATCTCC 191
Qy 952 GTAATCTGTCCAGCTGTAAACAGGTCCATCGAAAGCTCTGTTCGAAACAGCTCTCT 1011
Db 190 GTTGTCTGCGCGGTTGCAACACGCGCGCTCGCTAGCTCGTGGCGAACACAGCTCCT 131
Qy 1012 TGGGTCTCAACCATTTGGCGAGCAGCTGTGATCGAAGATTTCCCGCTGTGTCAGATTA 1071
Db 130 TGGGTCTCAACCATCGGAGCAGGAACGCTCGATAGAAATTTCCCGGGGTTGTTAGGTTA 71
Qy 1072 GCCAACGGAAAGTTCTTCTATGGAGAGTCAATGTATCCGGGAAAGGTATAAGAAATGCC 1131
Db 70 GCCAACGGGAGTACTCTACGGAGAGTCGTTGTACCCAGGGAAGGGATAGAGGCGA 11
Qy 1132 GGGAGAGAG 1140
Db 10 GAGAGAGTG 2

RESULT 5
BH59290/c
LOCUS BH59290 635 bp DNA linear GSS 15-DEC-2001
DEFINITION BOGWR58TR BOGW Brassica oleracea genomic clone BOGWR58, genomic
survey sequence.
ACCESSION BH59290
VERSION BH59290.1 GI:17851736
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eutrosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 635)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOGWR58TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..635
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOGWR58"
/clone_lib="BOGW"
/notes="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 127 a 198 c 162 g 148 t
ORIGIN
Query Match 18.8%; Score 438.2; DB 28; Length 635;
Best Local Similarity 80.6%; Pred. No. 1.2e-95;
Matches 512; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
Qy 635 AATACATTTCCGCAAGAGATTCACGGGACACGGGACTCACCGCTCACCGCTCACAGTGGTG 694
Db 635 AATACATATCGGACGAGACTCCACGGGCCACGGGACCCACACCGCTCACTGCTCGCG 576
Qy 695 GATCCTCTGTTTCGATCGGCAATGTTCTTGGCAATGAGCTGTGTGGCTCGTGGGATGG 754
Db 575 GCTCCTCTGTTTCGACGGCGAGCGTTCTCGTAAACGAGCGGGGTAGCCCGGGATGG 516
Qy 755 CTCCTGGAGCTCAATTTGCACTATAAAGTCTGTGGTTCATGTTGTGTATACAGTCTG 814
Db 515 CGCCAGAGCTCAGCTGCGGTTTACAAAGTCTGCTGTTCAACGGCTGTATACAGTCTAG 456
Qy 815 ACATTTAGACGCTATAGATGTAGCGATTCAGATAAAGTCGATGTTCTTTCGTTTCCC 874
Db 455 ACATTTCTGGCGGCGATAGACGTAGCGGATACAGACAAAGTCGACGTTCTATCCTCTTCCC 396
Qy 875 TTGCGGCTTTCCCTATTTCCTTTGTATGATGACCAATTCGCCATTGGAACATTCGAGCCA 934
Db 395 TAGCGGTTTCCCTATCCCTTTGATGATGACACTATCGCCATTGTTACCTTCCGCGCCA 336
Qy 935 TGGAAACCGGTATATCTGTAAATCTGTGACGTGGTAAACACCGGTCCAATCGAAAGCTGTG 994
Db 335 CGGAACATGGAATCTCCGTTGTCGCGGGGTGCAACACGCGCTCGCTAGCTCGG 276
Qy 995 TTGCAAAACAGCTCTCTTGGGTCTCAACCATTTGCGGCGACGCTGTGTATCGAAGATTC 1054
Db 275 TGGGAAACACAGCTCTCTTGGGTCTCAACCATTCGAGCAGGAACGCTCGATAGAAAATTC 216
Qy 1055 CCGCTGTGCTCAGATTAGCCAAACGGAAGCTTCTCTATGGAGAGTCAATGTTATCCGGAA 1114
Db 215 CGGGGTTGTTAGGTTAGCCAAACGGAAGCTATCTACGGAGAGTCTGTTGTACCCAGGA 156
Qy 1115 AAGGTATAAAGATGCGCGGAGAGAGGTTGAGGTGATTTACGTACAGGAGGAGATAAAG 1174
Db 155 AAGGATAAAGAGGCTGAGAGAGGTTGGAGATTGCTACGTCGCGGAGGAGACAAAG 96
Qy 1175 GAAGTGATTTCTGTTTCAGAGGGTCACTTCCAGAGAGAAAATCCGAGGCAAAATGTTGA 1234
Db 95 GAACCGAGTTTGTCTTCAGAGGCTCGCTTCCGAGAGAGAGTCTCCCAAGGCAAGATGGTGA 36
```

[illegible][illegible]

QY 1292 GAGGAGTTGCAATGATCTTTAGCAATACAGAGATCAACCAAGAAGAGATCTTATTGACG 1351
 DB 520 GCAGCGTCGGATGATCTTGGCTACCAACGAGATTAAACCAAGAGAGAGATCTTGTAGACG 461
 QY 1352 TTCACTCTTACAGCTACATGATTGGTTTACACTAGTCAGTCCTTCTGAAGGCTTATG 1411
 DB 460 TTCAATTTGATACCAAGCTACCTGATTGGTTTACAGAGATCTGTTGTTTGAAGAGTTACG 401
 QY 1412 TTAATCCAGGTTGAACCAAGCGCGGATATTTTGGTGTAGCGGTGATTGGGAGT 1471
 DB 400 TGAGGGACACCGGTAAGACCGAAAGCTAGGTTAATTTTCGGCGGACGGTGAATGGAGGT 341
 QY 1472 CACGAGCACCGGAGGTGGCTCAGCTTTTCAGCTCGAGGACCGAGTTTATGCAATCTTCGA 1531
 DB 340 CGAGAGGGCTTGAAGTGGCGCACTTCTCAGGAGAGACCGAGTTTGGCTAACCTTCGG 281
 QY 1532 TACTAAACCGGATATGATTCCTCGGGAGTCAATATCATTTGCGGCTTGGCTCATAATC 1591
 DB 280 TGCTTAAACCGGATTTGATCGCGCAGGTGTCAATATTATAGTGTGCTTGGCTCAGAAATC 221
 QY 1592 TAGACCAACCGGACTTCTTATGATTCAGAGAGGTTAACTTCACTGTAAATGTCAGGAA 1651
 DB 220 TTGACCGAGCGGGCTTCTTATGATTCGAGGAGGTTAACTTCACTGTGATGTCGGGA 161
 QY 1652 CTTCAATGCTTGTCCACATGTTAGCGGAATCACTGCTTATCCGGTCTGCATACCCGA 1711
 DB 160 CTTGCGATGCTTGTCCGACGCTGAGTGGAAATCACCGCTCTGATCCGGTCTGCTATCCGA 101
 QY 1712 ACTGGTCTCAGCTGCAATCAATCCGCAATGATGATCAACAGCGGATTTGTACGATCGTC 1771
 DB 100 GCTGGTCTCAGCTGCTATCAGATCAGGATGATGACGCGGCTGATTTGTATGATCGGA 41
 QY 1772 AAGGAAAGCGATAAGGATGTTAACAACACCGCGGTGT 1811
 DB 40 GAGGAAAGAGATTAGGACCGGGACAAACCGCGGGAGT 1

RESULT 8
 BH601308/c
 LOCUS BOHSA89TF BOHS Brassica oleracea genomic clone BOHSA89, genomic
 survey sequence.

ACCESSION BH601308
 VERSION BH601308.1 GI:17853754
 GSS.
 SOURCE Brassica oleracea
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 803)
 Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 Other_GSSs: BOHSA89TR
 Contact: Chris Town
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 1..803
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T01000D3"
 /db_xref="taxon:3712"
 /clone="BOHSA89"
 /clone_lib="BOHS"
 /note="Vector: phos1; Site_1: BstXI; 2-3 kb sheared

BASE COUNT 245 a 223 c 121 g 214 t
 ORIGIN genomic DNA inserted into phos1 using BstXI linkers"

Query Match 17.6%; Score 409.2; DB 28; Length 803;
 Best Local Similarity 80.6%; Pred. No. 1.4e-88;
 Matches 493; Conservative 0; Mismatches 113; Indels 6; Gaps 1;

QY 1723 GCTCAATCAAAATCCGCAATTCAGTACAAACAGCGGATTTTACGATCGTCAAGGAAAGCG 1782
 DB 803 GCTGCTATCATGATCAGGATGATGACGCGCTGATTTGATGATCGGAGGAAAGAG 744
 QY 1783 ATAAAGATGTTAAACAAACAGCGGCTGTTTTCGATTTGGAGAGGCGATGTAAATCCG 1842
 DB 743 ATTAGGACCGGGACAAACCGCGGGAGTGTGTTGCTATTGGAGCAGGCAATGTAAATCCG 684
 QY 1843 CAAAGCGGATAACCCGGGATTTGTTTACAACATTCACAGTGGATACATACCTTAC 1902
 DB 683 GTTAAGCGGATTAACCCGGGGTGGTTTACAACATTCACCGGTTGATTCATTCGTTAC 624
 QY 1903 CTCTGCACTCTTGGATTCACAGATCAGATATTTTACCAATCACTCATAGAAACGTCAGC 1962
 DB 623 TTATGCACTATGGGTTTACTAGATCGATATTTTACGATCACTCACAGGACGTTAGC 564
 QY 1963 TSCAATGGAATATTGCGGAAACCCCGGTTTATGCTCAATTAACCGCTCGATAGCGTG 2022
 DB 563 TCGGTCGGTACTGAGGAGAGCCCGGTTTATGCTTAAATATCCGTTATTCGGTT 504
 QY 2023 ATTTTCAAAAGTGCAAGACTACGAGATGATCACAAGGCTGTCTACTACGTTGGAGT 2082
 DB 503 ATTTTACGCTGGAGGACTAAGGAGATGATCAGAGGCTGTGACTAAACGTTGGAGT 444
 QY 2083 CCTAACGCTATATCACTCAGTGAATGTCAAGGCTCCAGAGGGATCAAAGTTATTGTCAAT 2142
 DB 443 CCTAACGCTGTACACTGTGAATGTGAAGCTCCTATGGGATTAATGTGATTTGAAG 384
 QY 2143 CCTAAGAGACTTGTGTTCAAAACAGCTGATCAGAGCTGAGCTATAGAGTATGATGTTGTA 2202
 DB 383 CCTAAGAGCTTGTGTTTATGCTACGTTGATCAACAGTTGAGCTATAGAGTTGTTGTG 324
 QY 2203 TTGAAGAAGAAAACAGAGAGG-----GAAGTGCTAGCTTTTGCACAGGCGAGTTG 2256
 DB 323 TTGAAGAAGGAGAGAGAGAGAGAGAGGCTGCTGCTACCTTTGCGGATGACAGATTG 264
 QY 2257 ACTTGGTCACTCTCATAATCTGATGAGGAGTTAGAGTCCAATCTCTGTAAACCTTG 2316
 DB 263 ACTTGGTCACTCTCGTATTCGATGCGAGTTAGGAGTCCAATCTCTGTACCTTG 204
 QY 2317 AAGACTAATCGA 2328
 DB 203 AAGAATCATTTGA 192

RESULT 9
 BH714875/c
 LOCUS BOMLB06TF BO_2_3_KB Brassica oleracea genomic clone BOMLB06,
 genomic survey sequence.

ACCESSION BH714875
 VERSION BH714875.1 GI:18809000
 GSS.
 SOURCE Brassica oleracea
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 527)
 Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 Other_GSSs: BOMLB06TR
 Contact: Chris Town
 TIGR

REFERENCE
 AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
 TITLE Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished
 COMMENT Other_GSSs: BOMLB06TR
 Contact: Chris Town
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.

FEATURES

Location/Qualifiers

1..527
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOML806"
 /clone_lib="BO_2_3_KB"
 /note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHO51 using BstXI linkers"

BASE COUNT 155 a 157 c 88 g 127 t

ORIGIN

Query Match 15.3%; Score 356.2; DB 28; Length 527;
 Best Local Similarity 81.2%; Pred. No. 1e-75;
 Matches 428; Conservative 0; Mismatches 93; Indels 6; Gaps 1;
 QY 1804 GCCGGTGTTCGGATGTGAGCAGGCGATGTAATCCGAAAGCGATAAACCCGGA 1863
 Db 527 GCGGGAGTGTTCGTATTGGAGCAGGCGATGTAATCCGGTTAAGCGATTAAACCGGGG 468
 QY 1864 TTGGTTTACACATTCACACAGGAGTATACATTAACCTCTGCACTCTGGATTACAA 1923
 Db 467 TTGGTTTACACATTCACACCGGTGATTACATTCCTTACCTATGCACTATTGGTTTACT 408
 QY 1924 AGATCAGATATTTAGCAATCACTATAGAACGTCAGTGTCAATGCAATATTGCGGAA 1983
 Db 407 AGATCGATATTTAGCGATCACTACAGAAAGTGTAGCTGCGGTCTAGTGGAGAG 348
 QY 1984 AACCCGGTTTTAGTCTCAATTAACCGTCGATAGCGGTGATTTTCAAGGTGCAAGAT 2043
 Db 347 AGCCCGGTTTTAGTCTTAATTAATTCGCTATTTCGGTTATTTTGTAGAGTGGAGGCT 288
 QY 2044 ACGGAGATGATCACAGGCGTGTCACTAAGTTGGGAGTCTTAACCTCGATATACTCAGTG 2103
 Db 287 AAGGAGATGATCACAGGCGTGTCACTAAGTTGGGAGTCTTAACCTCGGTGTACACTGTG 228
 QY 2104 AATGTCAGGCTCCAGAGGGGATCAAGATTATTTGTCATCTCAAGAGACTTGTTTCAAA 2163
 Db 227 AATGTTAAGGCTCTTATGGGATTAATGTAATTTAAGCCTAAGAGGCTTGTTTACT 168
 QY 2164 CACGTGATCAGACGCTAGCTATAGAGTATGGTTTGTATTGAAGAAGAAACACAGAGGA 2223
 Db 167 CACGTTGATCAACGTTGAGCTATAGAGTTGGTTTGTGTTGAAGAAGGAGACACAGA 108
 QY 2224 GG-----GAAGGTGGTAGCTTTGCAAGGGGAGTGTGACTTTGGGTCAACTCTCATAT 2277
 Db 107 GGAGAGAAGGTGGGTGTACTTTGCGGATGGACAGTGTGACTTTGGGTCAACTCTCGTGAT 48
 QY 2278 CTGATGACGCGAGTTAGAGTCCATCTCTGTAACTTTGAGACTAA 2324
 Db 47 TCGATGACGCGAGTTAGAGTCCATCTCTGTGACCTTGAAGAATCA 1

RESULT 10

BH793726
 LOCUS BH793726 767 bp DNA linear GSS 02-APR-2002
 DEFINITION ME MBa0001K22f Manihot esculenta Manihot esculenta genomic clone
 accession BH793726
 VERSION BH793726.1 GI:19891014
 KEYWORDS GSS.
 SOURCE Manihot esculenta (cassava)
 ORGANISM Manihot esculenta
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

REFERENCE 1 (bases 1 to 767)
 AUTHORS Tonkins,J.P., Pregene,M., Main,D., Goicoechea,J.L., Blackmon,B.,
 Atkins,M., Tohne,J. and Wing,R.A.
 TITLE New Genomic Resources for Cassava (Manihot esculenta): Development
 of a Deep-Coverage BAC Library and Preliminary STC Analysis
 JOURNAL Unpublished
 COMMENT Contact: Tonkins J
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
 Tel: 864 656 6419
 Fax: 864 656 4293
 Email: jtmkns@clemson.edu
 Total High Quality bases = 313
 Seq primer: TAATACGACTCACTATAGG
 Class: BAC ends
 High quality sequence start: 77
 High quality sequence stop: 546.

FEATURES

source

1..767
 /organism="Manihot esculenta"
 /mol_type="genomic DNA"
 /strain="MECW72"
 /db_xref="taxon:3983"
 /clone="ME_MBa0001K22f"
 /issue_type="Leaf"
 /lab_host="E. coli"
 /clone_lib="Manihot esculenta"
 /note="Vector: pCUGIBAC-1; Site 1: HindIII; Site 2: NotI;
 For more details on library preparation and sequence
 analysis see
 http://www.genome.clemson.edu/projects/etc/cassava/ME_MBa
 To order clones from this library see
 http://www.genome.clemson.edu/orders "

BASE COUNT 233 a 183 c 153 g 186 t 12 others
 ORIGIN

Query Match 14.0%; Score 326.8; DB 28; Length 767;
 Best Local Similarity 68.4%; Pred. No. 1.6e-66;
 Matches 473; Conservative 0; Mismatches 217; Indels 2; Gaps 2;
 QY 1400 TGAGGCTTATGTTAATGCGACCGTGAAACCAAGGCGCGGATAATTTTGGTGGTACGG 1459
 Db 70 TAAATCTTACATTAACCTTACAAGTAAACCAAGAGCTCGAATCATTTTCAAGGGAACGTG 129
 QY 1460 TGATGGGAGCTACGAGCAGCGGAGTGTCTAGTTTTCAGCTCGAGGACCGAGTTTGTAG 1519
 Db 130 TAAATCGGAAATCCAGAGCACCAGAGTAGCTCAGTTTTCAGCTAGAGGACCAAGTTTAA 189
 QY 1520 CCAATCTCTGATATAAACCAGGATATGATTGCTCCGGAGTCAATATCATTTCCGGCTT 1579
 Db 190 CTACCCCTTCAATCTCTCAACCCGATGTATTGCACTGGAGTAAACATCATTCGAGCTT 249
 QY 1580 GGCCTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAAGAAGAGTTAACTTCACTG 1639
 Db 250 GGCCTCAAAACTTAGGCCCACTGGTCTTCAGAGAATCCAAAGAAGAGTGAACCTTCACAG 309
 QY 1640 TAATGTAGGAACCTTCAATGTCTTGTCCACATGTTAGGGAATCACTGCTTATCCGGT 1699
 Db 310 TAAATGTAGGAACATCCATGGCTTGTCTCATGTGAGTGAATTTGCTGCCCTGATCCGCT 369
 QY 1700 CTGCATACCCGAACCTGGTCTCCAGTGCATCAATCAATCGCATTTGATGACAAACAGCGGATT 1759
 Db 370 CAGCTCATCCCGGATGGACACTCTGTCAGTTAATATGCAATATGACTACTGCAGATA 429
 QY 1760 TGTACGATCGTCAAGGGAAGCGATAAAGGATGTTAAACAAACCCAGCCGGTGTGTTGGGA 1819
 Db 430 TAACTGACCATCTCGAAATCCAAATATTGGATGTTGACAAACACGACGACTCTTTGCAA 489
 QY 1820 TTGAGGACGAGG-GCATGTGAATCCGCAAAAGCGGATAAACCCGGGATTGGTTTACACATT 1878
 Db 490 TTGAGGACGCGCATATACACCCAGAGAGCCATCAATCCGAGATTGATATATGATATC 549

QY 1879 CAACCAAGTGGATTACATAACTTACCTCTGCACTCTTGATTCACCAAGATCAGATATTTTA 1938
 DB 550 AGCCCCAGAGAGATGTCACCCACCTATGNCACCTAGATACANAGATCAGAAATTTTC 609
 QY 1939 GCAATCACTCATGAAGAGTGTAGCTCAATGGAATATTCGGAAACCCCGGTTTATG 1998
 DB 610 ACAATCACACAGGAATGTGAGCTGNCATGAACTATGATGATGAATGAAGGTTNCAGC 669
 QY 1999 CTCATTTACCGTCGATGCCGTGATTTTCAACGCTGGCAAGATACGAGATGATCACA 2058
 DB 670 CTCATATNATCTTNCATTTCTGATNTAAAGCG-GGGATGACNNAGTATGATGCCA 728
 QY 2059 AGCGGTGTCACTAACGTTGGAGTCTCACTC 2090
 DB 729 AGAGCACTTACCAATGTGGTATCCTTAAATC 760

RESULT 11

BF052773 568 bp mRNA linear EST 07-MAR-2003
 LOCUS EST438003 potato leaves and petioles Solanum tuberosum cDNA clone
 DEFINITION CSTB32P4 5' sequence, mRNA sequence.

ACCESSION BF052773
 VERSION BF052773.1 GI:10806669

KEYWORDS EST.

SOURCE Solanum tuberosum (potato)

ORGANISM

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Solanum.

AUTHORS

van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,
 Utterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Roming
 ,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.

TITLE Generation of ESTs from potato leaves and petioles

JOURNAL Unpublished

COMMENT Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato-array@tigr.org

This clone can be obtained from the University of Arizona Genomics

Institute. Orders can be made through URL:

http://genome.arizona.edu/orders/.

FEATURES

source

1..568

/organism="Solanum tuberosum"

/mol_type="mRNA"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="CSTB32P4"

/tissue_type="leaflets and petioles"

/dev_stage="8 weeks old plants"

/lab_hosts="SOLR"

/clone_lib="potato leaves and petioles"

/note="vector: pBluescript SK(-); Site 1: EcoRI; Site 2:

XhoI; Tissue was supplied by Dr. Fry (Cornell University).

Leaflets and petioles were isolated from 8 week old

greenhouse grown plants. The plants were watered and

fertilized freely. The tissue was immediately frozen in

liquid nitrogen."

BASE COUNT 176 a 123 c 133 g 136 t

ORIGIN

Query Match 12.7%; Score 296.8; DB 10; Length 568;

Best Local Similarity 70.4%; Pred. No. 3e-61;

Matches 397; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 1205 CAAGAGAGAAATCCGAGGCAAAATGGTGATTTGTATCCGGAGTCAATGGAGATCGG 1264

DB 4 CAAAAGCACAAATCCGAGGCAAAATGGTGATTTGTATCCGGAGTCAATGGAGGCGAG 63

QY 1265 AGAAAGGAGAGCGGGTTAAAGAGAGCTGGAGAGTTGCAATGATCTTAGCCATACAGAGA 1324

DB 64 AAAAAGCCAGGTTGTGAAGGAGCGAGGTGCTGCCATGATCTTAGCAATACAGCA 123
 QY 1325 TCAACCAAGAAGAAGATTCTATTGAGCTTCATCTTTACCAGCTACATTGATTGGTTACA 1384
 DB 124 TAAATATGGAGGAAGATTCGGTTGATGTCATGTCCTCCAGCAACGTTGATTGGCTCG 183
 QY 1385 CTGAGTCAGTCCTCTTCCTGAAGGCTTATGTTAATGCCACGGTGAAACCAAAAGCGCGGATAA 1444
 DB 184 ATGAATCAATTCAATTACAAAACCTACTGAACCTCAACAAAAGACCAACAGCTCGATTCA 243
 QY 1445 TTTTGTGTGTACGGTCAATTGGGAGGTCCAGCAGCAGCGGAGGTGGCTCAGTTTTCAGTTC 1504
 DB 244 TATTGGAGGACCGTAATAGGAAAGTCTAGAGCACCTGCAGTAGTCTCAGTTTTCTGTCAA 303
 QY 1505 GAGGACCGAGTTTAGCCAATCTTCGATACTAAACCGGATATGATGTCTCCGGGAGTCA 1564
 DB 304 GGGGCCCAAGCTATACGTATCTCAAACTGATTTGATGTTCTCCAGGGGTAA 363
 QY 1565 ATATCATTTGGCGTTGGCTTCAAAAATCTAGGACCAACCGGAGCTTCCTTATGATCAAGAA 1624
 DB 364 ACATAATTGGCGTTGGCCACAAAACCTTAAGCCCCAGTGTCTTCTCGAAGATTACGAA 423
 QY 1625 GAGTTAATCTCACTGTAATGTCCAGGAATCTCAATGTCTTCTCCACATGTTAGCGGATCA 1684
 DB 424 GAGTAAATTTCACTGTTATGTCCAGGACCTCAATGGCATGTCTCTATGTAAGTGAATTG 483
 QY 1685 CTGCTCTTATCCGTCGTGCATACCCGAACCTGGTCTCCAGCTGCAATCAAAATCCGCATTGA 1744
 DB 484 CCGATTGTCTCCATTCAGTCTATCTTAAATGGACTCCAGCAGCAATAAGATCCGCATAA 543
 QY 1745 TGACAACAGCGGATTTTGTACGATC 1768
 DB 544 TGACCACTGCAGATACAGCTGATC 567

RESULT 12

BH601319

LOCUS BH601319

DEFINITION BOHSA89TR BOHS Brassica oleracea genomic clone BOHSA89, genomic

survey sequence.

ACCESSION BH601319

VERSION BH601319.1 GI:17853765

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Brassica.

AUTHORS 1 (bases 1 to 772)

TITLE Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.

JOURNAL Whole genome shotgun sequencing of Brassica oleracea

COMMENT Unpublished

Other GSSs: BOHSA89TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..772

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TOL000DH3"

/db_xref="taxon:3712"

/clone="BOHSA89"

/note="vector: pBstXI; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pBstXI using BstXI linkers"

BASE COUNT	243 a	207 c	105 g	217 t
ORIGIN				
Query Match	11.7%	Score 273;	DB 28;	Length 772;
Best Local Similarity	80.9%	Pred. No. 1.9e-55;		
Matches 318;	Conservative	0;	Mismatches 75;	Indels 0;
Gaps				
QY	1	ATGGAACCCAAACCTTTCTTCTGTGCAATATCTTTCTTCTATTTTGTCTTCTTCTGCTCA	60	
DB	380	ATGGAACCTAAGACTCTCTTTGTCTTCACTATCTTTCTCTCTTTGTCTTCTTCTCTCTCA	439	
QY	61	GAGATCCTGCAGAACGAGACITACATTGTTTCAGCTTTCATCTCTAATAGCGAAACCGCTAA	120	
DB	440	GAAACCTTAAGAAACAGACTTACGTCATTACGCTTCACCTTAATAGCCCAATACGCTAA	499	
QY	121	ACCTTTGCGCTCAAAGTTTGATGGCATCTTTCTTTCTTCTCAAGAAGCGGTTTAGTGTT	180	
DB	500	GCCTTTCTTCAAAGCTTGATGGCACTTTCTTTCTTCAAGAGCTGTTCTAGGTATT	559	
QY	181	GAAGAAAGAGAGAGAGAGCTTTCTTCGACTTCTTCTATCTCTATGGCTCTCGAATTGAA	240	
DB	560	GAAGAAAGAAACGAGAGACCTTCTACAAGGATTCTTACTTCTACGCTCTGCGATTGAA	619	
QY	241	GGATTTGCTGCTCAGTTGACTGAATCAGAACCCGAGATCTGAGATATTCACCTGAAGTT	300	
DB	620	GGATTTCTCAGCTCAGTTAAACCGAATCAGAACCCAAAACACTCAAGAACTTACCTGAAAGT	679	
QY	301	GTTCGAGTCAGACCTGACCATGTTCTTCAGGTTCAAACCACTTACTTCTTACAGTTTCTTG	360	
DB	680	GTTCGAGTGAAGACCTGACCATGTTCTTCCAGTCAAACCACTTACTTCTTACAGTTCTTA	739	
QY	361	GCAGCTCGACGGTTTTGGAACTCCGGTGATGG	393	
DB	740	GGACTCAACGGTCCAGGACCTTCCAGCGTATGG	772	
RESULT 13				
CC340521/c				
LOCUS	CGRA039TV	ZM	0.7	1.5 KB
DEFINITION	789 bp DNA linear GSS 16-MAY-2003			
ACCESSION	genomic survey sequence.			
VERSION	CC340521			
KEYWORDS	CC340521.1	GI:30809927		
SOURCE	GSS.			
ORGANISM	Zea mays			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.			
AUTHORS	Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Numborg, A., Robbins, D., and Lakey, N.			
TITLE	Consortium for Maize Genomics			
JOURNAL	Unpublished			
COMMENT	Contact: Cathy Whitelaw			
	TIGR			
	9712 Medical Center Drive, Rockville, MD 20850, USA			
	Tel: 301-838-5843			
	Fax: 301-838-0208			
	Email: whitelaw@tigr.org			
	Seq primer: TF			
	Class: sheared ends.			
FEATURES	Location/Qualifiers			
source	1..789			
	/organism="Zea mays"			
	/mol_type="genomic DNA"			
	/strain="B73"			
	/db_xref="taxon:4577"			
	/clone="ZM0374H05"			
	/clone_lib="ZM_0.7_1.5_KB"			
	/note="Vector: pBGSK; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"			

```

Ehrhartoideae; Oryzae; Oryza.
1 (bases 1 to 750)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GCTATTAGGTGACACTATAG
Class: BAC ends
High quality sequence stop: 339.
Location/Qualifiers
1. .750
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSUNB008017r"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/notes="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa.
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
```

BASE COUNT 128 a 254 c 256 g 110 t 2 others

ORIGIN

Query Match 10.8%; Score 252; DB 28; Length 750;
Best Local Similarity 60.3%; Pred. No. 2.5e-50;
Matches 449; Conservative 0; Mismatches 292; Indels 3; Gaps 2;

QY 1321 GAGATCAACAGAGAGAGATTCTATTGAGCTTCATCTTACAGCTACATTGTTGCT 1380
DB 2 GAGATAACAGGAGAGAGACTCGTCGATGTCATGTCCTCGCTGAGCTATCGGC 61

QY 1381 TACACTGAGTCAGTCCTTCTTGAGGCTTATGTTAATGCCACGGTGAAACCAAGGCGCG 1440
DB 62 TACAGGAGGCGCTCGAGCTCAAGACTAGTCAGTCCAGCGCGCGCGGTGGCGAGG 121

QY 1441 ATATTTTGTGTAGCGGTGATTGGAGTGTACGAGCACCGGAGTGGCTAGTTTTCA 1500
DB 122 ATAGTGTTCGGCGGAGCGGATTTGGCGGGACGCGCGCGCGGTGGCTGTCTCG 181

QY 1501 GCTGAGGACCGATTTAGCCATCTTCGATCTAAACCGGATATGATTCGCGGA 1560
DB 182 GCGCGCGGCGGAGCCCTGACGAAACCCCTCGGTGCTCAAGCCGACGTGGTCGCCCGCGC 241

QY 1561 GTCAATATCATTTGGGCTTGCGCTCAAAATCTAGGACCAACCGGACTTCCTTATGATTC 1620
DB 242 GTGAACATCATCGCGGTGGCGCGCAACTGGGACCCGTGGGGCTGGAGGCGGAGCC 301

```

QY 1621 AGAAGAGTTAACTCACTGTAATGTGAGGAACCTTCAATGTCTGTGCCATGTTAGCGGA 1680
DB 302 CGCGCGTCCGACTTCACGGTGTCTCTCGGAGACATCAATGCGTGCCTGCCGACGTGAGCGC 361
QY 1681 ATACTGCTCTTATCCGGTCTGCATACCCGAACTGCTCTCCAGCTGCAATCAATCGCA 1740
DB 362 ATCGCGCGCTCATCAGGTTCGGCGACCCCGTCTGTGAGCCCGCGATGGTGGCGTCCGCG 421
QY 1741 TTGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAGACGATAAGAGTGTAAACAA 1800
DB 422 ATCATGACGACCGCGGACGTGACGACCGGAGGAGGAGCCGATCATGAGCGGACGCGC 481
QY 1801 CCA--GCCGGTGTGTTTGGGATTGGAGCGAGGGCGATGTAATCCGAAAAAGCGATAAAC 1858
DB 482 GCAAGGCCGAGCGGTACGCCATCGCGCGCGGCGACGTGAANCCGCGAGGCGGTGCACC 541
QY 1859 CGGGATTGGTTTACAACATTCACACAGTGGATTAACATTAACCTTACCTCTGCACCTTTGGAT 1918
DB 542 CGGGCTCTGCTACGACATCGACCGCGCGCGGACCTACCTACCCACCTCTTTGCACCTTCGGT 601
QY 1919 TCACAAGATCAGATATTTTAGCAATCACTCATAGAAGCGTGAGC-TGCAATGAATATG 1977
DB 602 ACACCCACATGGAGATCTTTAAGATCACCCACGCTTGGTCAACTTGCACCGCGGTGCTT 661
QY 1978 CGGAAAAACCGGGTTTTACTCTCAATTACCCGTCGATAGCCGCTGATTTTCAAACGTGGC 2037
DB 562 GAGCGGAGCGCGCTTACGCTCACTACCCATCATTANTGTTGGCGGTTTACGAGGACGAC 721
QY 2038 AAGACTACGAGATGATCAAGG 2061
DB 722 ACCACGTTGGCGGGCTCCCAAG 745

RESULT 15
BH720365 713 bp DNA linear GSS 20-FEB-2002
LOCUS BOHSY82TF BO 2.3 KB Brassica oleracea genomic clone BOHSY82,
DEFINITION genomic survey sequence.
ACCESSION BH720365
VERSION BH720365.1 GI:18921108
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 713)
AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other GSSs: BOHSY82TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: ctown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: Tg
Class: sheared ends.
Location/Qualifiers
1. .713
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHSY82"
/clone_lib="BO_2.3 KB"
/notes="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOSt using BstXI linkers"
BASE COUNT 143 a 190 c 203 g 177 t
ORIGIN
```

Search completed: February 3, 2004, 00:23:53
Job time : 3287 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:08:17 ; Search time 21 Seconds
(without alignments)

3549.082 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEKPFPLCIIFLLFCSSSS.....NGHNLQVRSPISVTLXTN 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4018	100.0	775	T00962	hypothetical prote
2	1742	43.4	754	A84473	probable serine pr
3	1679	41.8	775	T07172	subtilisin-like pr
4	1666.5	41.5	757	JC7519	subtilisin-like se
5	1640.5	40.8	746	S52770	subtilisin-like pr
6	1611	40.1	766	T07171	subtilisin-like pr
7	1589	39.5	784	T05768	subtilisin-like pr
8	1572.5	39.1	754	T06577	subtilisin-like pr
9	1562	38.9	756	G86150	F22M8.3 protein -
10	1542	38.4	754	T06579	subtilisin-like pr
11	1534.5	38.2	745	T07184	subtilisin-like pr
12	1513.5	37.7	747	T06580	subtilisin-like pr
13	1490	37.1	745	JC6119	subtilisin-like pr
14	1445.5	36.0	758	T51335	subtilisin-like pr
15	1282	31.9	761	S52769	subtilisin-like pr
16	1272	31.7	774	A86454	hypothetical prote
17	1271.5	31.6	756	T04187	subtilisin-like pr
18	1236	30.8	777	C86454	hypothetical prote
19	1221	30.4	766	T48389	cucumis-like pro
20	1220.5	30.4	761	T07169	subtilisin-like pr
21	1219.5	30.4	779	C86335	hypothetical prote
22	1216.5	30.3	763	B86454	hypothetical prote
23	1215.5	30.3	775	T04189	subtilisin-like pr
24	1213	30.2	769	D86335	T20H2.6 protein -
25	1201	29.9	736	JC7518	subtilisin-like se
26	1196	29.8	803	T04190	subtilisin-like pr
27	1190.5	29.6	747	T04188	subtilisin-like pr
28	1188.5	29.6	779	T07170	subtilisin-like pr
29	1187.5	29.6	706	T01351	subtilisin-like pr

30	1180.5	29.4	736	2	T12963	subtilisin homolog
31	1176	29.3	739	2	T12964	subtilisin homolog
32	1173	29.2	765	2	T04186	subtilisin-like pr
33	1172.5	29.2	731	1	A55800	cucumis (EC 3.4.
34	1145.5	28.5	746	2	T06017	subtilisin-like pr
35	1143	28.4	753	2	B96687	subtilisin-like pr
36	1134.5	28.2	758	2	A96887	hypothetical prote
37	1117.5	27.8	779	2	T14845	antifreeze-like pr
38	1117	27.8	772	2	T05838	subtilisin-like pr
39	1101.5	27.4	755	2	T48553	subtilisin-like pr
40	1076.5	26.8	718	2	T05840	subtilisin-like pr
41	1070	26.6	769	2	T05939	subtilisin-like pr
42	1067	26.6	1736	2	T05174	hypothetical prote
43	989	24.6	687	2	D85165	cucumis importe
44	985	24.5	734	2	D86454	F9L11.14 F9L11.14
45	907.5	22.6	815	2	T00538	probable serine pr

ALIGNMENTS

RESULT 1

T00962
hypothetical protein F20D22.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T00962
R:Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, G.; Oji, K.; Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence.
A:Reference number: Z14214
A:Accession: T00962
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-775 <VYS>
A:Cross-references: EMBL:AC002411; NID:G2570223; PID:G3142298; GSPOB:GN00059; ATSP:F20L
C:Genetics:
A:Gene: ATSP:F20D22.12
A:Map position: 1
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homolog

Query Match	100.0%	Score	4018	DB	2	Length	775
Best Local Similarity	100.0%	Pred. No.	2.5e-270				
Matches	775	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
QY	1	MEKPFPLCIIFLLFCSSSS	IIKQYIYVQLHPNSETAKTFASKFDWHL	SFLQEA	VLGV	60	
DB	1	MEKPFPLCIIFLLFCSSSS	IIKQYIYVQLHPNSETAKTFASKFDWHL	SFLQEA	VLGV	60	
QY	61	EEEEEPSSRLLYSGAIGFAAQLT	ESAEILRYSPVVAVR	PDVQLQV	QITTSYKFL	120	
DB	61	EEEEEPSSRLLYSGAIGFAAQLT	ESAEILRYSPVVAVR	PDVQLQV	QITTSYKFL	120	
QY	121	GLDGFSGNWSKSRFGQGTII	IGVLDGVNPFSPD	TGMPSIPR	KWKGICQEGESFSS	180	
DB	121	GLDGFSGNWSKSRFGQGTII	IGVLDGVNPFSPD	TGMPSIPR	KWKGICQEGESFSS	180	
QY	181	SSCNKLTIGARFFIRGHRVANS	PEESPMPREYISARDST	CHGTHTAST	VGSSVS	240	
DB	181	SSCNKLTIGARFFIRGHRVANS	PEESPMPREYISARDST	CHGTHTAST	VGSSVS	240	
QY	241	LGNGAGVARGMAPGAHIAV	KVCFWNGCYSSDILAAID	VAIQDVLS	ISLGGFP	300	
DB	241	LGNGAGVARGMAPGAHIAV	KVCFWNGCYSSDILAAID	VAIQDVLS	ISLGGFP	300	
QY	301	DDTATCTFRAMERGISV	ICAAGNNGPIESSVANT	APWSTIGAGT	IDRRP	360	
DB	301	DDTATCTFRAMERGISV	ICAAGNNGPIESSVANT	APWSTIGAGT	IDRRP	360	
QY	361	KLLYGESLYPGKGNAGRE	VEVIYVGGKSGSEFCL	RGLSPREIR	GKWCIDR	420	
DB	361	KLLYGESLYPGKGNAGRE	VEVIYVGGKSGSEFCL	RGLSPREIR	GKWCIDR	420	

Qy	421	SEKGEAVKEAGGVAMILANT	INQEDSDIVHLLPATLIGT	YESVLLKAYVNAVTPKAR	480	
Db	421	SEKGEAVKEAGGVAMILANT	INQEDSDIVHLLPATLIGT	YESVLLKAYVNAVTPKAR	480	
Qy	481	IIIFGGTVIGRSRAPEVAQ	FARGPSLANPSILKPDMI	APGVNIAAQPQNLTGTPYDS	540	
Db	481	IIIFGGTVIGRSRAPEVAQ	FARGPSLANPSILKPDMI	APGVNIAAQPQNLTGTPYDS	540	
Qy	541	RRVNFVTMSGTSMSCPHV	SGITALLRSAYPNWSPA	AKSALMTTADLYDRQKAI	KGNNK 600	
Db	541	RRVNFVTMSGTSMSCPHV	SGITALLRSAYPNWSPA	AKSALMTTADLYDRQKAI	KGNNK 600	
Qy	601	PAGVFAIGAGHVNPOKAI	NPGLYVNTQPDYIYTL	CTLGTFRSDILAI	THKNVSCNGILR 660	
Db	601	PAGVFAIGAGHVNPOKAI	NPGLYVNTQPDYIYTL	CTLGTFRSDILAI	THKNVSCNGILR 660	
Qy	661	KNPFGSLNYPISIAVIF	RKGKTKTEIMTRRV	TNVGSPNSISYVNV	KAPEGIKVI	VNPKRLVF 720
Db	661	KNPFGSLNYPISIAVIF	RKGKTKTEIMTRRV	TNVGSPNSISYVNV	KAPEGIKVI	VNPKRLVF 720
Qy	721	KHVDQTLTSYRWFWFL	KKNKRGKGVASFAQ	GLTWNHNLNQVRSP	ISVTLTKN 775	
Db	721	KHVDQTLTSYRWFWFL	KKNKRGKGVASFAQ	GLTWNHNLNQVRSP	ISVTLTKN 775	

RESULT 2
 A84473
 probable serine proteinase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: A84473
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koc, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eises, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: A84473
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-754 <STO>
 A:Cross-references: GB:AE002093; NID:g4006827; PIDN:AAC95169.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g05920
 A:Map position: 2
 C:Superfamily: subtilisin-like proteinase ag12; subtilisin homology

Query Match	43.4%;	Score 1742;	DB 2;	Length 754;
Best Local Similarity	48.4%;	Pred. No. 1.3e-112;		
Matches	370;	Conservative 118;	Mismatches 237;	Indels 40; Gaps 16;
Qy	11	IFLFPCSSSSILQKQYIVOLHNSSETAKTFASKFDWHLSFLQEAVLGVVEEERSSR	70	
Db	18	LFLHLHTA-----KKTIIIRVN-HSKPESFLTHDWTYSQLNS-----ESS	59	
Qy	71	LLYSGSAIEGFAAQLTSEAE-IIRYSPEVAVRDPHVQLQVQTVTSYKFLGLDG-FGNS	128	
Db	60	LLYYTTSFHGFSAYLDSSTEADSLSSNSILDIFEDPLYLTHTRTFEFLGLNSEFGVH	119	
Qy	129	GVWSKSRFGQGTIGVLDTGVWPSPSPDDTGMPSIPRKWKIGICQEESEFSSSSCNKLI	188	
Db	120	DLGSSS---NGVIIGLDTGVWPSPSRPDDTDMPEIPSKWKGECEGSDPDKLCKNKLI	176	
Qy	189	GARFPIRHRVANSPEESPNNPREYISARDSTGTHGTHTASTVGSSSVSMANVLNGAGVA	248	
Db	177	GARSFSGKQFQWASGGGFSK--RESVSPRDVDHGHTTSTTAAGSAVRNASFGLYAAGTA	234	
Qy	249	RGMAPGAHIAVYKVCWNGCYSSDILAAIDVAIOKDVDVLSLGLGPPILPYDDTTAIGT	308	
Db	235	RGMATRVARVYKVCWSTGCTGSDILANDRAILDGVLDVLSLGLGGSAPYRRTIAlGA	294	
Qy	309	FRAMERGISVICAAGNNGPIESSVANTAPWVSTTIGAGTLDRRFPAVVRLANGKILYGESL	368	

Db	295	FSAMERGVPFVSCAGNSGPTRASVAVPVMWTVGAGTLDPRDFAPANLGNQKRLTGVSL	354
Qy	369	YPKGIGKNAGREVEVIVTGDGSEFCLGSLPREBIRCKWICDRGVNGRSEKGEAVK	428
Db	355	YSGVGM--GTPLELVTKNGNSSNNCLPSSLDSSIVRGKI VVCDGVNARVEKGAVR	412
Qy	429	EAGGVAMILANTEINOBEDSIDVHLLPATLIGYTESVLLKAYNATVKPKARIIFGGTVI	488
Db	413	DAGGLGIMANTAAASGBELVADSHLLPAIAVGKKTGLLREYVKDSKPTALLVFKGTVL	472
Qy	489	GRSRAPVEAOFARGPSLANPILKPDMIAPGVNIIAAMPQNLPQTGLPYDSSRRVNTVM	548
Db	473	DVKPSPVAAAFSRGPNTVTPILKPDVIGPGVNILAGNSDAIGTGLDKDSRETQFNIM	532
Qy	549	SGTSMSCPHVSGITALIRSAYPNKSAAIKSALMTTADLYDRQKAKDGG--NKPAGVEA	606
Db	533	SGTSMSCPHISGLAGLLKAAHPWSPSAIKSALMTTAYVLDTNTNAPLHDAADNSLPNYA	592
Qy	607	IGAGHVNPQAINPGLVNIOPVDYIPLYCLTGLFTRSDILAITHK--NVSCNGILRKNPGF	665
Db	593	HGSHVDQKALSPGLYDITSTREYIFPLCSLDYTVDHI VAI VKRPSVNCCKKF--SDPG-	650
Qy	666	SLNYPSTAVIFKRGKTTMITRRVNTVNGSPNSIYSVNVKAPEGIKVTPKRLVFKHVDQ	725
Db	651	QLNYPSPSVLF--GGKRVVRYTRVNTVNGAASSVYKVTVNGAPSVGISVYKPSKLSFKSVGE	709
Qy	726	TLSYRVWVFIKKXNRGKGKVASFAQ--GQLTVWNSHNLMOVRSPDIS	769
Db	710	KKRYTVTFVSKK---GYSMTNKAEPFGSITWNSNPOH---EVRSPVA	748

RESULT 3

T07172

subtilisin-like proteinase (EC 3.4.21.-) 2 - tomato

C:Species: Lycopersicon esculentum (tomato)

C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000

C:Accession: T07172

R:Meichery, J.; Anshein, N.; Schaller, A.

A:Submitted to the EMBL Data Library, May 1998

A:Description: The gene family of subtilisin-like proteases in tomato.

A:Reference number: Z15771

A:Accession: T07172

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-775 <MEI>

A:Cross-references: EMBL:AJ006379; PIDN:CAA07000.1

A:Experimental source: cultivar VFw8

```

C:Genetics:
A:Gene: sbt2
C:superfamily: subtilisin-like proteinase ag12; subtilisin homology
C:keywords: hydrolyase; serine proteinase

      Query Match      41.8%; Score 1679; DB 2; Length 775;
      Best_Local Similarity 46.0%; Pred. No. 3,2e-108;
      Matches 359; Conservative 125; Mismatches 261; Indels 36; Gaps 16

QY      6 FFLCI--IFLFFCSSSEILQKQYIVOLHPNSTAK--TFASKFDWHLAFLEQAVLGV 60
Db      11 FFSVYCLAINLAKSPNT---KKTYYIQM---DKWAKPDEVFDVQVSYSLVKSVLPSF 63
QY      61 DEEEF--FPSRLLYSYGATGFAAQLTSEAEILRYSPEVAVRPHVLQVQYTSYK 118
Db      64 TEVEKGTGDEERILYSYQTAFFGVAQAQSEEVKKLQERNGLVAFPEIKYQLHTIRSP 123
QY      119 FLGLDGFNSGWSKSRFGQGTIIIGVLDTGWVPESPSPDDTGMPSIPRKWKIGIQEGESF 178
Db      124 FLGLDREUSSKLWADRLSDHNVIIVGLDTGWIPESPSPDSCMTSPVSHWKGVCGTCGRF 183
QY      179 SSSSNRKLIGARFFIRGHVANSPEEPGNMPREVISARDSTGHGHTHTASTVGGSSVMA 238
Db      184 EKXHKCKIIVGARVFRFGVEAASG---KINRGEFKSARDQDGHGHTHTAGTVAGSVVRA 240

```



```

QY 239 NVLGNAGVARGMAPGAHIAVYKVCWFNGCYSSDILAAIDVAIODKVDVLSLSLGGPPIP 298
DB 241 NLLGAYGTARGMAPGARVAAYKVCWVGCFSSDILSAVDOAVADGVNILLSLGGVSS 300
QY 299 LYDDTIAICTFRAMERGISVCAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLA 358
DB 301 YNRSLSLAARGAMEKGVFVSCSAGNGGPDPSLITNVSPWITTVGASTMDRDPATVELG 360
QY 359 NGKLLYGBSLYPGKIRKNAGREVEVYVYTGDKG---SEFCRLGSLPREIRGQWVICDR 415
DB 361 TGIKIVTGASLYKGRMNLSTQOYPLIYLGSSNLSMPSSLCDGTLDKASVAGKIVICDR 420
QY 416 GNGRSEKGEAVKEAGGVAMILANTEINOEBSDIDVHLLPATLIGYTEVLLKAYVNAVY 475
DB 421 GISPRVQGVYKGEAGGVGMILTNTAANGELVADSHLLPAVAVGEGEGRRAIKLYA-AGR 479
QY 476 KPKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTG 535
DB 480 SATATLRLGTLGIRPSPVVAARSRGPNFLSLLEILKPDMPAGVNVILAGNTGALGPSS 539
QY 536 LPYDSRRNVFTVMSGTSMSCPHVSGITALLIRSAYPNWSPPAAIKSALMTTADLYDRQKAI 595
DB 540 LPIDQRRNPNLSTGSMSCPHVSGIAALLKARHPDMSPPAAIKSALMTTAYVHDNTYKSL 599
QY 596 KQGNK--PAGVFAIGAGHVNPKAINPGLVNIQPVYIYVLCGLGTRSDILAI-THKN 652
DB 600 KDASVTPTSDYDHGAGHVNPRKAVDPLGIYDIGAQDYFEFLCTQELSPSQLVVFGKFSN 659
QY 653 VSCGILRNKPFSLNPSIAVIFKRGKTTMIT--RRVTVNGSPNSIYSNVYKAPGK 710
DB 660 RTCHSL-ANPG-DLNPAPISAVPPEKTKLSMLTLHRTVTVNGSPIYNYHVYVSFAFGAV 717
QY 711 VIVNPKRLVFKHVDQTLRYVWFVLLKKNGRGKVASFAQGOLTVWNSHNLMOVRSPISV 770
DB 718 VKVEPERLNFTSKQKLSYKTKTVSRQ---KAPF--GSLIWKDG---THKVRSPIAI 769
QY 771 T 771
DB 770 T 770

```

RESULT 4

```

JC7519
N:Arabidopsis-like serine proteinase (EC 3.4.21.-) - Arabidopsis thaliana
N:Alternate names: subtilase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7519
R:Yamagata, H.; Usugi, M.; Saka, K.; Iwasaki, T.; Aizono, Y.
Biosci. Biotechnol. Biochem. 64, 1947-1957, 2000
A:Title: Molecular cloning and characterization of a cDNA and a gene for subtilisin-like
A:Reference number: JC7518; MUID:20506592; PMID:11055401
A:Accession: JC7519
A:Molecule type: DNA
A:Residues: 1-757 <YAM>
A:Cross-references: GB:AF065639
C:Comment: This enzyme is organ-specific and is involved in the specific proteolytic eye
C:Genetics:
A:Gene: asp48
C:Superfamily: subtilisin-like proteinase ag12; subtilisin homology
C:Keywords: hydrolase; serine proteinase

```

```

Query Match 41.5%; Score 1666.5; DB 2; Length 757;
Best Local Similarity 46.4%; Pred. NO. 2.3e-107;
Matches 362; Conservative 122; Mismatches 246; Indels 51; Gaps 18;

```

```

QY 6 FFLCIIFFLFCSSSEILQKQYIVQLHPNSETAKT-PASKFDWHLFLQBAVLGVREEE 64
DB 11 FFL-LLCGFCRVSSSSDQQTIVHM-----AKQMPSSFDLHSNWDSSLRISD-- 61
QY 65 EEPSSRLIYSGATEGPAALQTESEAEILRYSPEWAVRPHVLQVQTYSYKFLGLDG 124
DB 62 ---SAELLYTYENALHGFSRLTQEEADSLMTQGVISVLPEHRYELHTTTPFLGLDE 118

```

```

QY 125 -----FGNSGVWKSFRFGGTLIIGVLDTGVPWPESPFDGTGMPSTIPRKWKIGICQEGSEFS 179
DB 119 HTADLFFPAGSYSS-----DVVVGVLDTGVWPESPSSYSDGFGPIPSWKMGKCEAGTNT 172
QY 180 SSSCNKRLIGARFFIRGHRVANSP--EESPNNPREVISARDSTGHTGHTTASTVGGSSVMA 238
DB 173 ASLCNRKLLIGARFPARGVESTMGPIDES-----KESRSPRDDDGHGHTHTSSAAGSVVGA 228
QY 239 NVLGNAGVARGMAPGAHIAVYKVCWFNGCYSSDILAAIDVAIODKVDVLSLSLGGPPIP 298
DB 229 SLGVSAGTARGMAPRARVAVYKVCWLGCFSSDILAAIDKAIADNVANVLSLGGGMSD 288
QY 299 LYDDTIAICTFRAMERGISVCAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLA 358
DB 289 YRVDGVAIGAFAMERGIIVSCSAGNAGPSSSLSNVAPWITTVGAGTLDLDRFPALALIG 348
QY 359 NGKLLYGBSLYPGKIRKNAGREVEVYVYTGDKG--EFCLGSLPREIRGQWVICDRG 416
DB 349 NGKNFTGSLFKGEALPD--KLLPFIYAGNASNATNGMLCMTGLIPBKVKGIWMCDRG 406
QY 417 VNGRSEKGEAVKEAGGVAMILANTEINOEBSDIDVHLLPATLIGYTEVLLKAYVNAVYK 476
DB 407 INARVQKGDVVKAGGVGMILANTANGELVADAHLLPATTVGEGKAGDIIRHYVTTDPN 466
QY 477 PKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGL 536
DB 467 PTASISILGTVVGVKPSPVAAFPSSRGPNSTPNILKPDLIAPGVNIIAAMPQNLGPTGL 526
QY 537 PYDSRRNVFTVMSGTSMSCPHVSGITALLIRSAYPNWSPPAAIKSALMTTADLYDRQKAIK 596
DB 527 ASDSRREVENIISGTSMSCPHVSGLAALLKSVHPEWSPAAIRSAALMTTAYKTYKDGKPLL 586
QY 597 D--GNKPAGVPAIGAGHVNPKAINPGLVNIQPVYIYVLCGLGTRSDILAIHKNVS 654
DB 587 DIATGKPSPPDHGAGHVSPTTATNPGLIYDUTTEDYLGFLCALNYTSPQIRSVRRNT 646
QY 655 CNGILRNKPFSLNPSIAVIFKRGKTTMIT--RRVTVNGSPNSIYSNVYK--PEGIK 710
DB 647 CD----PSKSYSVADLNPSPFAVVD--GVGAKYKTRTVTSVGGAGT--YSVKYVSETTVK 700
QY 711 VIVNPKRLVFKHVDQTLRYVWFVLLKKNGRGKVASFAQGOLTVWNSHNLMOVRSPISV 770
DB 701 ISVEPAVLNFKEANERKSYTVTFTVDSKPSG---SNSFGSIEWSDGKHV---VGSVAI 754
QY 771 T 771
DB 755 S 755

```

RESULT 5

S52770

```

subtilisin-like proteinase (EC 3.4.21.-), nodule-specific - Arabidopsis thaliana (frag
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 22-Jun-1999
C:Accession: S52770
R:Ribeiro, A.; Akkermans, A.D.L.; van Kammen, A.; Bisseling, T.; Pawlowski, K.
submitted to the EMBL Data Library, March 1995
A:Description: A nodule-specific gene encoding a subtilisin-like protease is involved :
A:Reference number: S52769
A:Accession: S52770
A:Molecule type: mRNA
A:Residues: 1-746 <AIB>
A:Cross-references: EMBL:X85974; NID:g757533; PIDN:CAAS9963.1; PID:g757534
C:Function:
A:Description: Involved in early stages of actinorhizal nodule development
C:Superfamily: subtilisin-like proteinase ag12; subtilisin homology
C:Keywords: hydrolase; serine proteinase

```

```

Query Match 40.8%; Score 1640.5; DB 2; Length 746;
Best Local Similarity 46.0%; Pred. NO. 1.4e-105;
Matches 359; Conservative 122; Mismatches 248; Indels 51; Gaps 19;

```

```

Qy 6 FLCLIFLLFCSSSEIILOKQTYIVOLHPNSETAKTFASKFDWHLSPLOEAVLGVVEEBE 55
Db 2 FFL-LLCLGFCFHVSSSSDOQTYIVHM-AKSQT----PSSFDLHNSNYSUSSRISD--- 52
Qy 66 EFSRLLYSGAIEGFAQLTESAEILRSPWAVRCHVLQVOTVSYKFKLGLD- 124
Db 53 --SAELLYYENAIHGFTRIQBEADSLMTQPGVISVLPEHRYELHTTPTPLFLGLD 110
Qy 125 ----FNGSGWKSRRFGOGTIIGVLDTCWPSPSFDDTCWPSIPRKNWKGIOEGESFSS 180
Db 111 TADLFPAGSYS-----DVVVGVLDTGWPESKSYDEGFTPSSWKGCEAGNFATA 164
Qy 181 SSCNRKLIARFIRGHRVANSP-EESPNMREVISIARDSTGHGTHTASTVGGSSYSMAN 239
Db 165 SLICNRKLIARFARGYESTMGPIDES-----KESRSPRDDGHTHTSSTAAGSVVEGAS 220
Qy 240 VILNGAGVARGMAPGAHIAVYKVCWFNGCYSSDILAAIDVAIOKVDVLSLGGPPIPL 299
Db 221 LIGYASGTARGMLHA--LAVYKVCWLGCFSSDILAAIDKAIADNVNLSLSCGGWSY 278
Qy 300 YDDTIAIGTFRAMERGISYICAAANGNGPIESSVANTAPWSTTIGAGTLDRRPPAVVRLAN 359
Db 279 YRDGVAIGAFAAMERGILVCSAGNAGPSSSLSNVAFWITTVGAGTLDRDPFALLAIGN 338
Qy 360 GKLLYGESILPGKGIKNAGREVEVIYVGGDKGS--EFCLRGSLPREETRGKVI CDRGV 417
Db 339 GKNFTCVSLFKGEALPD--KLIPFTIYAGNASNATNGNLCMTGTLPEKVGKTIWMDCGI 396
Qy 418 NGRSEKGEAVKAGGVAMILANTEINQBEDSIDIYHLLPATLILGYTESVLKAVNATVP 477
Db 397 NARVQGDVVKAAGVGVMILANTAANGBELVADHLLPATTVBEKAGDIIHRVYTTDPNP 456
Qy 478 KARIIPGGTVIGRSRAPEVAOFSARGPSLANPSILKPDMIAPGVNIIAAMPONLPGTLP 537
Db 457 TASISILGTVGVGKPSVVAAFSSRGSPNSITPENILKPDLIAPGVNIIAAMTGAAGPTGLA 516
Qy 538 YDSRRNVFTVMGTSMSCPHYSGITALLIRISAYPNWSPAAIKSALMTTADLYDRQKAIKD 597
Db 517 SDSRRVEFNIIISGTSMSCPHYSGLAALIKSHVPEWSPAAIRSAALMTTAYKTYDGRPLD 576
Qy 598 --GNKPGAVFATGAGHVNPOKAINPGLVYNIQPDVITYLCTLGFQTESDILATHTKNVSC 655
Db 577 IATGKPSPTDFHGAGHVSEPTATNPGLYIDUTTEDYLGFLCALMYTSPQIRSVRRNYTC 636
Qy 656 NGILRXNPGFS--LNPYSIAVIFRGRKHTEMIITRRVTNVGSPNSIYSVNVKA-PEGIKV 711
Db 637 D----PSKYSYVADLNYPSFANVD-GAGAKYKRTVTSVGGAGT-YSVKVTSEITGVKI 690
Qy 712 IYNPKELVFKHVQDTLSYRVTVLKKKNRGKGVASFAGQQLTWNVNSHNLNQRVSRPISVT 771
Db 691 SYEPAVLNFKEANEKSYVTFVTVDSSKPSG---SNSPGFIESWDGKHV--VGSFVAIS 744

```

RESULTS

subtilisin-like proteinase (EC 3.4.21.-) 1 - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T07171

R;MeiChtry, J.; Anthrein, N.; Schaller, A.
submitted to the EMBL Data Library, May 1998
A;Description: The gene family of subtilisin-like proteases in tomato.
A;Reference number: Z15771

A;Accession: J07171
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-766 <ME1>
A;Cross-references: EMBL:AJ006378; PIDDN:CA006999.1
A;Experimental source: cultivar VFWS

RESULT 7

RESULT
T05768

subtilisin-like proteinase (EC 3.4.21.-) - *Arabidopsis thaliana*
N; Alternate names: protein M4E13.40

C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 13-Aug-1999
C/Accession: T05768
R/Bavan, M.; Purnelle, B.; Boutry, M.; Goffeau, A.; Heuvelink, J.; Mewes, H.
submitted to the Protein Sequence Database, April 1998
A/Reference number: Z15451
A/Accession: T05768

A:Molecule type: DNA
 A:Residues: 1-764 <BEV>
 A:Cross-references: EMBL:ALJ022023
 A:Experimental source: cultivar Columbia; BAC clone M4E13
 C:Genetics:
 A:Map position: 4
 A:Note: M4E13.40
 C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
 C:Keywords: hydrolase; serine proteinase

Query Match 39.5%; Score 1589; DB 2; Length 764;
 Best Local Similarity 44.0%; Pred. No. 5.5e-102;
 Matches 342; Conservative 127; Mismatches 268; Indels 40; Gaps 16;

QY 7 FLCIFLLFCSSSEILQKQTVVQLHPNSETAKTFASKFDWHLSPLOEAVLGVBEERE 66
 DB 11 FLSPFFISFASQA---AKTFPRIDGGS-MPSIPTHYHWY-----STFAER- 55
 QY 67 PSRLYSYSAIEGFAAQLTESEAILRYSPEVAVRPHVLQVOTVYSYKFLGDGFG 126
 DB 56 --SRIVHVHTVPHGFSAVVTPDEADLNKHPAVLAVFEDRRRLHTTSPQFLGLQ--N 111
 QY 127 NSGVWKSRRGQGTIIIVLDTGVWPSPSPDDTCMPISIPRKWKICQEGESSESCNRK 186
 DB 112 QKGLWSESDYGSVDIIGVFTGTPWERRSFSDNLGPIPKRWGVCESGARFSRNCNRK 171
 QY 187 LIGARPPIRGHRVANSPEESPNMPREVIGARDSTGHGTHTASTVGGSSVSMANVLGAG 246
 DB 172 IIGARFPAKQQA---VIGINTVFELSPRDADGHGTHSTTAAGHAFKASMGYASG 229
 QY 247 VARGMAPGAHIAVYKVCW-FNGCYSSDILAAIDVAIQDKVDVLSLSLG---GFPILYDD 302
 DB 230 VARGVAPKARIAAVKVCWDSGCLDSDILAAFDAAVRDGVVDVISIGGDDGITSPPYLD 289
 QY 303 TIAIGTFRAMEGIVSICAGNNGPIESSVANTAPWVSTIGAGTLDRRPAPVVRLANGKL 362
 DB 290 PIAAGSYGAASKIGFVSSSAGNEGPNMSVTNLAPWTVTGASTIDRNPADAILDGH 349
 QY 363 LYGESLYPGKCIKNAGREVEVIYV-TGGDKGSFECRLGSLPREIRGKVICDRGVNGRS 421
 DB 350 LRGVSLYAGVPLN--GRMFFVYVPGKSGMSSASLCMENTLDPKQVRGKIVICDRGSSPRV 407
 QY 422 EKEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGVTESVLLKAYVNATVVKPARI 481
 DB 408 AKGLVVKAGGVGMILANGSNEGVLGDHILIPACAVSNEGDRIKAVASHNPPIASI 467
 QY 482 IFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVN-IAAWPQNLGTGLPYDSR 541
 DB 468 DFRGTIVGKPAFVIASFSGRPNGLSPEILKPDLIAPGVNIIAAMTAVGFTGLPSDR 527
 QY 542 RVNFTVMGSMSCPHVSGITLIRSAYPNWSAPAAIKSALMTTADLYDRQGAIKDGN-- 599
 DB 528 KTEFNILSGTSMACPHVSGAALLKSAHPDWSPAVIRSAMMTTNLVDNSRSLIDESTG 587
 QY 600 KPAGVPAIGAGHVPQKAINPGLVYNIQPDVITYLCTLGFTRESDLAITHKNVSCNGIL 659
 DB 588 KSATPDYDGGHNLGAMNPGVLDITNDYITFLCSIGYGPKTIQVITRTPVRCPTTR 647
 QY 660 KKPFGFSLNPSIAVIF--KRGKTTETMIRRVNTNVGSPNSIYSVNVKAPGKIVNPK 716
 DB 648 KPSPG-NLNVPSITAVPTWRGLVSKVITRTATNVQAEAVYRARIESPGRGVTVTKPP 706
 QY 717 RLVPKHKVDQTLRYRVWFLVKKXN--RGKVASFAQQQLTVWNSHNLQMRVSPISVT 771
 DB 707 RLVFTSAVKRSYAVTVVTRNVWLGETCAVE--GSVTFWFDGGKHV--VRSPIVVT 759

RESULT 8
 T06577
 subtilisin-like proteinase (EC 3.4.21.-) - tomato
 N:Alternate names: proteinase p69c
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000

C:Accession: T06577
 R:Meichtry, J.; Amrhein, N.; Schaller, A.
 submitted to the EMBL Data Library, May 1998
 A:Description: The gene family of subtilisin-like proteases in tomato.
 A:Reference number: Z15771
 A:Accession: T06577
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-754 <MEI>
 A:Cross-references: EMBL:AJ005171; PIDN:CAA06412.1
 A:Experimental source: cultivar VF88
 C:Genetics:
 A:Gene: p69c
 C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
 C:Keywords: hydrolase; serine proteinase

Query Match 39.1%; Score 1572.5; DB 2; Length 754;
 Best Local Similarity 44.1%; Pred. No. 7.5e-101;
 Matches 346; Conservative 122; Mismatches 253; Indels 63; Gaps 19;

QY 7 FLCIFLLFCSSSEILQK--QTYIVQLH-PNS--ETAKTFASKFDWHLSPLOEAVLGV 60
 DB 3 FLKIFFVFIFCSPWPPIQSDFTFYIVHVESPSLITQSSFMWLESYVLSFLPETMSAI 62
 QY 61 EEEEEPSSRLLSYSGAIEGFAAQLTESEAILRYSPEVAVRPHVLQVQTYYSYKFL 120
 DB 63 SSSGNEEAAASIIYSYHNVTGFAARLTAEQVKEMEKHKGFVSAQKQRIILSLHTHTPSFL 122
 QY 121 GLDGFSGWWSKRFQGTIIIGVLDTGVPWPSPSPDDTCMPISIPRKWKICQEGESFSS 180
 DB 123 GLQ--QNKGVWKDSNYKGVIIGVLDGTIIPDPSPFSDVQMPSPPAKWKVC---KSNFT 177
 QY 181 SSCNRKILGARFFIRGHRVANSPEESPNMPREVISARDSTGHGTHTASTVGGSSVSMANV 240
 DB 178 NCKNKLIGARSYELGN-----ASPDNDGHGTHTASTAAGAPVKGANV 221
 QY 241 LGNGAGVARGMAPGAHIAVYKVCWFG--CYSSDILAAIDVAIQDKVDVLSLSLGFFIPL 299
 DB 222 HGNANGTAVGVAPLAHIAIKVCGDKCPCSDILAAMDAIIDDGVILSISLGSLSPL 281
 QY 300 YDDTIAIGTFRAMEGIVSICAGNNGPIESSVANTAPWVSTIGAGTLDRRPAPVVRLAN 359
 DB 282 YDETIALGAVSTTQRGILVSCAGNSGSPASVNSAPWITVGASTDLRKIKATVKLGN 341
 QY 360 GKLLYGSLIPGK-----GIKNAGREVEVIYVGGDKGSEFCLGSLPREIRG 408
 DB 342 GEFEFEGESAVHPKTSNATFTFLFDAKNKADPSETPY-----CRGSLTDPAIRG 391
 QY 409 KMWIC-DRGVNCRGEKGEAVKEAGVAMILANTEINOEDSIDVHLLPATLIGVTESVLL 467
 DB 392 KIVLCIAPGGVAVVQKQAVKAGGVGMIVNPSQYGVTKSADAHVLPALVVSADGTKI 451
 QY 468 KAYVNATVVKPARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAM 527
 DB 452 RAYTNSILNPVATITFGGTIIGDKNAPIVAAFSSRGENTASRGILKPDIIIGPVNIIAAM 511
 QY 528 PQLNGPTGLPYDSRRVNTVMGSMSCPHVSGITLIRSAYPNWSAPAAIKSALMTTADL 587
 DB 512 PTSVDGN---KNTKSTFNIIISGTSMSCPHLSGVAALLKSHPDWSPAVIKSAIMTTADT 567
 QY 588 YDRQGAIKDGN--KPAGVFAIGAGHVPQKAINPGLVYNIQPDVITYLCTLGFTSRDIL 646
 DB 568 LNLASSPILDBRLSPADIYAIAGHVPSPRANDEGLVYDTFFEDLYFLYCLGNTNSQVG 627
 QY 647 AITHKNVSCNGILRKNGFSLNPSIAVIFKRGKTTETMIRRVNTNVGSPNSIYSVNVKAP 706
 DB 628 KLLKKNKNCSEV-ESIPEAQLNYPFC-ISELGSTPQTFRTFTVNVGDAKSSYTVQIASP 685
 QY 707 EGKIVYNPKELVFKHVDQTLRYRVWFLVKKKNGKGVASFAQQQLTVWNSHNLQMRVRS 766
 DB 686 KGVVVKVKPKRLIFSELKQKLTQVTF--SKRTNSKSGGVF-EGFLKW-NSNK--YSVRS 739
 QY 767 PISV 770

```

Db      740 PIAV 743
||:|
||:|
602 AGVFAIGAGHVNPKAINPGLVNIQFVYITVLTGLTFRSDILAIHKNVSC--NGIL 559
589 ATAFAGAGNVDPTRAVDPLGVYDTSTVDYLNLYCSLNTYTSRILFLSGTNYTCASNAV 648
Qy      660 RKMPGSLNYPSTAVIFKRGKTTMI--TRRVNVSNSIYSVNVKAPGKIVNPKR 717
Db      649 -LSPG-DLNPSPFAVNLVNGANLKTVRYKRTVTVNGSPICEYVMHVEEPKVRPEPKV 706
Qy      718 LVFKHVDOTLSYRVWFVLKKNKGKVASPAQQLTWV-NSHNLMOVRSPISVT 771
Db      707 LKFOKAEERLSYTVTD-AEASNSSSSP--GVLWICDKYN---VRSPIAVT 754

RESULT 10
T06579
subtilisin-like proteinase (EC 3.4.21.-) p69e - tomato
C|Species: Lycopersicon esculentum (tomato)
C|Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C|Accession: T06579
R|Meichtry, J.; Amrhein, N.; Schaller, A.
A|Submitted to the EMBL Data Library, May 1998
A|Description: The gene family of subtilisin-like proteases in tomato.
A|Reference number: Z15771
A|Accession: T06579
A|Status: preliminary; translated from GE/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-754 <MEI>
A|Cross-references: EMBL:AJ005172; PIDN:CAA06413.1
A|Experimental source: cultivar VFWM
C|Genetics:
A|Gene: p69e
C|Superfamily: subtilisin-like proteinase agl2; subtilisin homology
C|Keywords: hydrolase; serine proteinase

Query Match      38.4%; Score 1542; DB 2; Length 754;
Best Local Similarity 43.4%; Pred. No. 9.7e-99;
Matches 341; Conservative 119; Mismatches 261; Indels 64; Gaps 19;

Qy      6 FLICIFLLFCSSSEILQK--QTYIYQLH-PNS--ETAKTFASKFDWHLFLQEAVLGV 60
Db      3 FLKILLVLIFCSFPWPTIQSGLEIYIVHVESPELSLSTQSFDTLDSYISFLPETTSAI 62
Qy      61 EEEEEPSSRLLYSGVNAIEGFAAQLTSEAEILRYSPVAVRFDHVLQVOTYSYKFL 120
Db      63 SSSGNEEAATWYIYHNVTGFAARLTASHVKEKKGFGVSAQKQILSLDTHTPSFL 122
Qy      121 GLDGFSGSVWKSFRFGQGTIIIGVLDTGVPWPEPSFDDTGMPSPRKKWKGICQGESEFSS 180
Db      123 GLQ--QNWGVWKNYKGVIIIGVLDITGLPDHPSFSDVGMPPPPAKWKGYC---ESNFT 177
Qy      181 SSCNKLIGARFFTRGRHVANSPEESNMPREYISARDSTGHTHTASTVCGSSVSNV 240
Db      178 NKCNNKLGARSYHLNG-----SPIDGDGHGHTTASTAAGAFVKGANV 221
Qy      241 LQNGAGVARGVAPGHAHIAVYKWFN--GCYSSDILAAIDVAIQKVDVLSLSLGFPIPL 299
Db      222 YGNANGTAGVAPLAHIAVYKVCSSDGGCSDSILAAIDVLDGVDLISISIGGSPNSL 281
Qy      300 YDDTIAIGTFAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDRFPVAVRLAN 359
Db      282 YDDPIALGAYSATARGVFCVCSAGNRPGLLASVGNAPWILTGVASTLDRKIKATVGLGN 341
Qy      360 GKLLYGESLYPK-----GKNAGREVEIYVYTGDKSEFCRLRSLPREIRG 408
Db      342 GEEFEGESAYRQTSNSTFTFLDAAKADPSETPY-----CRPGSLTDPVIRG 391
Qy      409 KWCIC--PRGVNGRSEKGEAVKEAGGVAMILANTFINGEEDSIDVHLLPATLIGVTSVL 466
Db      392 KIVLCACGVSS--VDKGVKWDAGGVGVINPNSQYGVTKSADAHVLPALDVSADGTR 450
Qy      467 LKAYVNAVTKPKARIIFGTVIGRSAPVAFQSFARGPSLANPSILKPDMIAPGVNIIAA 526
||:|
||:|

```

```

Db      740 PIAV 743
||:|
||:|
602 AGVFAIGAGHVNPKAINPGLVNIQFVYITVLTGLTFRSDILAIHKNVSC--NGIL 559
589 ATAFAGAGNVDPTRAVDPLGVYDTSTVDYLNLYCSLNTYTSRILFLSGTNYTCASNAV 648
Qy      660 RKMPGSLNYPSTAVIFKRGKTTMI--TRRVNVSNSIYSVNVKAPGKIVNPKR 717
Db      649 -LSPG-DLNPSPFAVNLVNGANLKTVRYKRTVTVNGSPICEYVMHVEEPKVRPEPKV 706
Qy      718 LVFKHVDOTLSYRVWFVLKKNKGKVASPAQQLTWV-NSHNLMOVRSPISVT 771
Db      707 LKFOKAEERLSYTVTD-AEASNSSSSP--GVLWICDKYN---VRSPIAVT 754

RESULT 9
G86150
F22M8.3 protein - Arabidopsis thaliana
C|Species: Arabidopsis thaliana (mouse-ear cress)
C|Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C|Accession: G86150
R|Theologian, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A|Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A|Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A|Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A|Reference number: A86141; MUID:21016719; PMID:11130712
A|Accession: G86150
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-756 <STO>
A|Cross-references: GB:AE005172; NID:G8570441; PIDN:AAF76468.1; GSPDB:GN00141
C|Genetics:
A|Map position: 1
C|Superfamily: subtilisin-like proteinase agl2; subtilisin homology

Query Match      38.9%; Score 1562; DB 2; Length 756;
Best Local Similarity 44.0%; Pred. No. 4e-100;
Matches 341; Conservative 130; Mismatches 268; Indels 36; Gaps 18;

Qy      7 FLICIFLLFCSSSEILQKTYIYQLHPNSETAKTFASKFDWHLFLQEAVLGVVEEBE 66
Db      6 FIVFFFLIFASNVUS-SRKQTYVHTVTTT-TKHIVTSFN---SLQTEINI---NDPDF 56
Qy      67 PSSRLLYSGAIEGFAAQLTSEAEILRYSPVAVRFDHVLQVOTYSYKFLGDLGFG 126
Db      57 SLPEIHYIENAMSGFATLTDQLDTVKNTKGFISAYPDELLSLHTTYSHEFLGLE-FG 115
Qy      127 NSGVWKSRRGQGTIIIGVLDTGVPWPEPSFDDTGMPSPRKKWKGICQGESESSCNRK 186
Db      116 -IGLWNETSLSSDVIIGLVDTGISPEHVSPRDTHTMTVPGRWGSCDEGNFSSSCNKK 174
Qy      187 LIGARFFTRGRHVANSPEESNMPREYISARDSTGHTHTASTVCGSSVSNVNLGNAG 246
Db      175 IIGASAFYKY--ESIVGKINETTDFRSTRDAQHGHTTASTAAGDIVPKANYFGQAKG 231
Qy      247 VARGVAPGAHTAVYKVCNFGCYSSDILAAIDVAIQKVDVLSLSLGGFPPIPLYDTTIAI 306
Db      232 LASGWRFTSRIAAYKACWALGCASTDVIADRAILDGVDVLSLSLGGSSRPFPVDPFIAI 291
Qy      307 GTTFAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDRFPVAVVLANGLLYGE 366
Db      292 AGFGAMQKNIFVCSAGNSGFTASTVNSGAPWLTVAASYTDRTPFAIVRIGNRKSIVGS 351
Qy      367 SLYPKGKIKNAGREVEIYVYTGDKSEFCRLRSLPREIRGKAVICDRGVNGRSEKGEA 426
Db      352 SLYKGSLSKUL--PLAFNRTAGEESGAVFCIRDSLKEELVEGKIVICLRASGRTAKGEE 409
Qy      427 VKEAGGVAMILANTFINGEEDSIDVHLLPATLIGYTESVLLKAYVNAVTKPKARIIFGGT 486
Db      410 VKESGGAAMLLVSTEAGEEELLADPHVLPVLSLFGSDGKTLNLYLAGAANATASVFRGT 469
Qy      487 VIGSRAPVAFQFARGPSLANPSILKPDMTAPGNIIAQPWNLGPTGLPYDSRRNVFT 546
Db      470 AYG-ATAPVAAFSRSGSVAGPEIAKPDIAAPGLNLAGSPFSSPILLSRSDPRRYQFN 528
Qy      547 VMSGTSMSCPHVSGITAIRSAYPNWSPAAIKSALMTTADLYDRQGAIKD-----GNKP 601
Db      529 IISGTSMACPHISGIAALIKSVHGDWSPAMIKSAIMTITARNRPIGDRGAAGAES 588

```

```

Db 451 IRAYTNSILNPVATITFOGTHIGDENAPIVAAFSSRGPNNTASGILKPDIIIGGVNIIAA 510
Qy 527 WPNLPGTGLPYDSRRVNTVMGTSMSCHPVSGITALLRSAYPNWSPAIAKSLMTTAD 586
Db 511 WPTSVNCGN-----KNTKSTENIIISGTSMSCHPVSGIILKSGHSPDWSPAIVKSAIMTTAD 566
Qy 587 LYDRQGAIAKDGN-KEAGVPAIGAGHVNPKALNPGLVNIIQPDVYITLCTLGFTSRDI 645
Db 567 TNLASSPIIDELKSLPADIIYAIAGHVNPSRANDPGLVYDTPTFEDYLPYLCGLNNTNSQV 626
Qy 646 LAITHKNVSCNGILRNKPNPSLNPSTIAVIFKRGKTEMITRRVTNVGSPNSIYSVNVKA 705
Db 627 GKLLKKNVCSV-ESIPERLAQLNYPSC-ISRIGSTPQTFQTFRTVTNVGDAKSYTVQIAS 684
Qy 706 PEGIKVINPKRIVKHVDOTLSYRVNVLKKNRGKGVASPAQGLTWNHNLNMQVR 765
Db 685 PKGVVWVKRKLIFSELKQKLYQVTF--SKRTNSKSGVF-EGFLKW-NSNK--YSVR 738
Qy 766 SPISV 770
Db 739 SPIAV 743

RESULT 11
T07184
subtilisin-like proteinase (EC 3.4.21.-) precursor P69B, pathogenesis-related - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-Jul-2000
C:Accession: T07184
R:Toranzo, P.; Conejero, V.; Vera, P.
J. Biol. Chem. 272, 14412-14419, 1997
A:Title: Identification of a new pathogen-induced member of the subtilisin-like processin
A:Reference number: Z15980; MUID:97306355; PMID:9162080
A:Accession: T07184
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-745 <TOR>
A:Cross-references: EMBL:Y10149; NID:G2230958; PIDN:CAA71234.1; PID:G2230959
A:Experimental source: cultivar Rutgers; leaf
C:Genetics:
A:Gene: P69B
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-115/Domain: signal sequence and amino-terminal propeptide #status predicted <PRO>
F:115-745/Product: subtilisin-like proteinase #status predicted <WAT>

Query Match 38.2%; Score 1534.5; DB 2; Length 745;
Best Local Similarity 43.3%; Pred. No. 3.2e-98;
Matches 339; Conservative 109; Mismatches 266; Indels 69; Gaps 17;

Qy 10 IIFLLFCCSSSEILQK--QTYIVQLH-PNS--ETAKTFASKFDWHLSPFQEAVLGVVEEE 64
Db 7 LLVFIFCSFQWPTIQSNLETVIVHVESPELVTTQSLLDLGSYYLSFLPKTATTISSG 66
Qy 65 EEPSSRLLYSGAIEGFAAQLTESAEILRYSPEVVAVRPDHLVQVQTYTSYKFLGLG 124
Db 67 NEENATWYIYHNVMGTGFAARLTAQVKENEKKGHGFVSACKQKILSLHTHPSPFLGLQ- 125
Qy 125 FNGSGVWKSRRFGQGTIIIGVLDTGVPSPSPDGTGMPSPRKGICQEGESFSSSCN 184
Db 126 -QNGVWVKDSYNGKGVIGVDTGILPDHPSFDVGMGPPPAKWKVC--ESNFTKCN 181
Qy 185 KLIIGARFFTRGHRVANSPEESNMPREYISARDSTGHGTHTASTVGGSSVSMANVLNG 244
Db 182 NKLIGARSYQLGNG-----SPIDSGHGHGTHTASTAAGAFVKGANYGNA 225
Qy 245 AGVARGMAPGAHTAVYKVCWFGNGCYSDILAAIDVAIQDKVDVLSLGLGFFPLVDDTI 304
Db 226 DGTAVGAPLAHTAIYKVNCSGSDVLAAMDALDGDVLSLSLGGPIPFHRDNI 285
Qy 305 AIGTFRAMERGISVICAAGNNGPIESSVANTAPWSTVIGAGTLDRFPAPVRLANGKLLY 364
Db 286 AIGAYSATERGILVSCSAGNSGSPFITAVNTAPKILTVGASTLDRKIKATVKLNGGEETE 345

```

```

Qy 365 GESLYPK-----GIKNAGREVEVIYVTGGDKSGSEFCFLRGLSPREEIRGKWVIC 413
Db 346 GESAYRPKISNATFTTDFDAKNAKDPSETPY-----CRRGSLTPAIRGKVLVC 395
Qy 414 DR-GVNGSRSEKGEAVKAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVN 472
Db 396 SALGHVANVKGQAVKDAGGVGMIIINPSQYGVTKSADAHVLPALVVSADGKIKILAYMN 455
Qy 473 ATVKPKARIIPGGVIVIGRSRAPEVAQPSARGPSLANPSILKPDMIAPGVNIIAAMPQNLG 532
Db 456 STSPVATIAFOGTIIIGDKNAPVAAFSRSGPRASPGILKPDIIIGPGANILAAWPTSD 515
Qy 533 PTGLPYDSRRVNTVMGTSMSCHPVSGITALLRSAYPNWSPAIAKSLMTTADLYDRQ 592
Db 516 DN-----KNTKSTENIIISGTSMSCHPVSGIILKSGHSPDWSPAIVKSAIMTTADTLNAN 571
Qy 593 KAIKDG-N-KEAGVPAIGAGHVNPKALNPGLVNIIQPDVYITLCTLGFTSRDIATHTK 651
Db 572 SPILDERLLPADIIYAIAGHVNPSRANDPGLVYDTPTFEDYLPYLCGLKYDQVGNLIQR 631
Qy 652 NVSCN---GILRNKPNPSLNPSTIAVIFKRGKTEMITRRVTNVGSPNSIYSVNVKAPEG 708
Db 632 RVNCEVKSIIEA---QLNVPSPS-IFGLGSTPQTYTRVTNVGDATSSYKVEVASPEG 686
Qy 709 IKVINPKRIVKHVDOTLSYRVNVLKKNRGKGVASPAQGLTWTY-NSHNLNMQVRSP 767
Db 687 VAIEVERSELNSELNQLTYQVTF---SKRTNSNPEVIEGFLKWTNSRHS-----VRSP 739
Qy 768 ISV 770
Db 740 IAV 742

RESULT 12
T06580
subtilisin-like proteinase (EC 3.4.21.-) p69f - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C:Accession: T06580
R:Meichtry, J.; Amrhein, N.; Schaller, A.
submitted to the EMBL Data Library, May 1998
A:Description: The gene family of subtilisin-like proteases in tomato.
A:Reference number: Z15771
A:Accession: T06580
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-747 <MEI>
A:Cross-references: EMBL:AJ005173; PIDN:CAA06414.1
A:Experimental source: cultivar VFW8
C:Genetics:
A:Gene: p69f
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
C:Keywords: hydrolase; serine proteinase

Query Match 37.7%; Score 1513.5; DB 2; Length 747;
Best Local Similarity 42.5%; Pred. No. 9e-97;
Matches 331; Conservative 129; Mismatches 267; Indels 51; Gaps 16;

Qy 6 FFLCIIFLLFCCSSSEILQK--QTYIVQLHNPNSSETAKTFASKFD--WHLSPFQEAVLGV 60
Db 3 FLKILLIFICSFLRPSIQSDLETVIVHVESPENQISTQSSLTDLSEYLSFLPKTTAI 62
Qy 61 EEEBEPSSRLLYSGAIEGFAAQLTESAEILRYSPEVVAVRPDHLVQVQTYTSYKFL 120
Db 63 SSSGDEEAASMIYSHNVNMGKFAARLTAQVKENEKKGHGFVSACKQKILSLHTHTSPFL 122
Qy 121 GLDGFSGVWKSRRFGQGTIIIGVLDTGVPSPSPDGTGMPSPRKGICQEGESFSS 180
Db 123 GLQ--QNGVWVKDSYNGKGVIGVDTGILPDHPSFDVGMGPPPAKWKVC--ESNFT 177
Qy 181 SSCNKRKTIGARFFTRGHRVANSPEESNMPREYISARDSTGHGTHTASTVGGSSVSMANV 240

```

```
Db 178 TKNNKLIGARSYQLONG-----SPIDNNGHGTHTAGTAAGAFVKGANI 221
Qy 241 LGNGAGVARGMAPAHAIVYKVCWFN-GCVSSDILAAIDVAIQDKVDLSLGLGFPPIPL 299
Db 222 FGNANGTAVGAPLAHAIVYKVCSSDGGSDSDILAMPAAIDDDGVDILSLSGSTKFF 281
Qy 300 YDDTIAIGTFRAMERGISVICAGNNGPIESSVANTAPWVSTIGAGTLDLRRFPAPVRLAN 359
Db 282 HDDGIALGTYSATERGIFVSASAGNSGSLGTVANEAPWILTVGASTHDKLKVTKLGN 341
Qy 360 GKLLYGESLPYKGIKNAGREVEVIYVGGDKGSE-----FCLRGLSPREIEGKVICDR 415
Db 342 SEEFEGESAYHP---KTSNSTFFPLDAGKNESDQSPFSPGSLNDPAIKGIKIVLCRL 398
Qy 416 GVN-GRSEKGEAVKEAGGVAMILANTEINQEDSIDVHLLPATLIGYTESVLLKAYNAT 474
Db 399 SISLLRVAQGSVDKAGGVNMLINEQEBGVTKSAEAHVLPALDVSNADGKILAYMNS 458
Qy 475 VKPKARIIFGGTIGVGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAPQNGLPT 534
Db 459 SNPVASTIFGTVIGDKNAPIVAFSRSRGSVASPGLKPDIIIGPGVNVLAAPMTSDVN 518
Qy 535 GLPYDSRRVNTVWGTSMSCPHVSGITLIRISAYPNWSPAALKSALMTTADLYDRQKA 594
Db 519 ----KNTKSTFNIVSGTSMSCPHLSGVAAALLKGAHPDWSPAALKSAMMTTADTVNLNSP 574
Qy 595 IKGNK-PAGVFAIGAGHVPQKAINPGLVYNIQPDYITYLCTLGTSTRDILAIHKV 653
Db 575 ILDBERLSADLEFAGAGHVPNSRASPGLVDYTPFDYIPYLCGLNTYREVGVKVLQKV 634
Qy 654 SCNGILRKNPGFSLNYPISIAVIFKRGKTTETMIRRVTVNGSPNSIYSVNVKAPGKIV 713
Db 635 NCSEV-KRIPEGQLNYPFSI--RLGSTPQTYRTTVNVGDAKSSYKVEIVSPKVVVKV 691
Qy 714 NPKRLVFKHVDQTLISYRVWFLKKNRGGKVASFAOQLTW-VNSHNLQVRVRSPIV 770
Db 692 EPSALNFSTLNQKLYQVIFT---KTNISTTSDVEGFLKWNNSRHS----VRSPIAV 742

RESULT 13
JC6119
subtilisin-like proteinase (EC 3.4.21.-) - tomato
N:Alternate names: calcium-activated endopeptidase; proteinase p69
C:Species: Lycopersicon esculentum (tomato)
C>Date: 23-Mar-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
C:Accession: JC6119
R:Tornero, P.; Conejero, V.; Vera, P.
Proc. Natl. Acad. Sci. U.S.A. 93, 6332-6337, 1996
A:Title: Primary structure and expression of a pathogen-induced protease (PR-P69) in tom
A:Reference number: JC6119; MUID:96270538; PMID:8692815
A:Accession: JC6119
A:Molecule type: mRNA
A:Residues: 1-745 <TOR>
A:Cross-references: EMBL:X95270; NID:q1524114; PIDN:CAA64566.1; PID:G1524115
C:Comment: This enzyme, a secreted calcium-activated endopeptidase, is a plant pathogene
fense against attacking pathogens.
C:Genetics:
A:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:137-546/Domain: subtilisin homology #status atypical <SBT>

Query Match 37.1%; Score 1490; DB 2; Length 745;
Best Local Similarity 42.5%; Pred. No. 3.8e-95;
Matches 337; Conservative 111; Mismatches 267; Indels 78; Gaps 19;

Qy 6 FFLCIIILLPCSSSEILQK--QTYIVQLHPNSETAKTPASKPD---WHLSFLQEAVLGV 60
Db 3 FLKILLVIFCSFWPTIQSNLETYLVHVSPPBSLSTQSSLDLDSYILSLFLPKTTAI 62
Qy 61 EEEBEPSSRLYSGVGAIEFAQLTSEAEILRYSFEVAVRPHVLOVQVTSYKFL 120
Db 63 SSSGNEEAATMIYSHNVMTGFAARLTAEQVKEKIHGFVSAQKQRTSLDITHTSSFL 122
```

```
Qy 121 GLDGFNGSGVWSKSRFGQTHIIGVLTGVWPBPSPFDDTGMPISIPRKWKIGICQEGSFSS 180
Db 123 GLQ--QNGWVKEDSNYKGVIIIGVITGILDPSPSDVGMPPPPAKWKGVGVC---ESNFT 177
Qy 181 SSCNKLIGARFTRGHRVANSPEESPMPREVIISARDSTGHGHTHTASTVGGSSVSMANV 240
Db 178 NCKNNKLIGARSYQLONG-----SPIDDDGHGHTHTASTAAGAFVNGANV 221
Qy 241 LGNGAGVARGMAPAHAIVYKVCWFNGCYSSDILAAIDVAIQDKVDLSL-L-GGFPPIPL 299
Db 222 FGNANGTAAAGVAPPAHAIVYKVCNSDGCADTVLAAMDAAIDGVDILSLSLGGGSSDF 281
Qy 300 YDDTIAIGTFRAMERGISVICAGNNGPIESSVANTAPWVSTIGAGTLDLRRFPAPVRLAN 359
Db 282 YSNPIALGAYSATERGILVSCSAGNNGPSTGSGVNEAPWILTVGASTQDRKLVKATVKLGN 341
Qy 360 GKLLYGESLPYKGIKNAGREVEVIYVGGDKGSEFCLRGLSPREIEGKVICDR 408
Db 342 REEFEGESAYRPKTSNSTFFALPDAGKNASDEFETPY-----CRSGSLTDPVIRG 391
Qy 409 KMWICDRGVN-GRSEKGEAVKEAGGVAMILANTEINQEDSIDVHLLPATLIGYTESVLL 467
Db 392 KIVICLAGGVPRVDKQAVKADAGVGMIINQORSVTKSADAHVIPALDISDADGTKI 451
Qy 468 KAYVNAVTKPKARIIFGGTIGVGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAW 527
Db 452 LAYNSTSNFVATITFGTIIIGDKNAPIVAFSRSRGSVASIGILKPDIIIGPGVNIILAAW 511
Qy 528 PONIUGPTGLPYDSRRVNTVWGTSMSCPHVSGITLIRISAYPNWSPAALKSALMTTADL 587
Db 512 PTSVDN----KNTKSTFNIIISGTSMSCPHLSGVAAALLKGAHPDWSPAALKSAMMTTADT 567
Qy 588 YDRQKAIKGN-KPAGVFAIGAGHVPQKAINPGLVYNIQPDYITYLCTLGTSTRDIL 646
Db 568 LNLANSFILDRLLPADIYALGAGHVPNSRANPGLVDYTPFDYIPYLCGLNTYREVGV 627
Qy 647 AITHKNVSCN---GILRKNPGFSLNYPISIAVIFKRGKTTETMIRRVTVNGSPNSIYSNV 703
Db 628 NLLQKVNCSKVEKSIIEA---QLNYPFS-IVDLGSTPQTYRTTVNVGDAKSSYKVEV 682
Qy 704 KAPGKIVYNPKLVFK---HVDQTLISYRVWFLKKNRGGKVASFAOQLTW-VNSHN 759
Db 683 ASPEAL-----PSKLTLRANPSSDKLYQVTF---SKTANSSNTEVIEGFLKWTNRHS 734
Qy 760 LMQRVRSPIVTL 772
Db 735 ----VRSPIALLL 743

RESULT 14
TS1335
subtilisin-like proteinase AIR3, auxin-induced [imported] - Arabidopsis thaliana (fragm
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Nov-2000
C:Accession: TS1335
R:Neuteboom, L.W.; Ng, J.M.Y.; Kuyper, M.; Clijdesdale, O.R.; Hooykaas, P.J.J.; van der
Plant Mol. Biol. 39, 273-287, 1999
A:Title: Isolation and characterization of cDNA clones corresponding with mRNAs that ac
A:Reference number: Z25377; MUID:99178779; PMID:10080694
A:Accession: TS1335
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-758 <NEU>
A:Cross-references: EMBL:AF055848; PIDN:AAC62611.1
A:Experimental source: root culture
A>Note: accumulates during auxin-induced lateral root formation
C:Genetics:
A:Gene: AIR3
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology

Query Match 36.0%; Score 1445.5; DB 2; Length 758;
Best Local Similarity 41.3%; Pred. No. 4.8e-92;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:07:12 ; Search time 17 Seconds
(without alignments)

2143.864 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEKPFPLCIIIFLLFCSSSS.....NSHNLQVRVRSIVTLKTN 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	430	10.7	806	1 SUBV_BACSU	P29141 bacillus su
2	271.5	6.8	382	1 SUBT_BACAM	P00782 bacillus am
3	264.5	6.6	361	1 ELYA_BACHD	P41363 bacillus ha
4	264.5	6.6	1902	1 P2P_IACPA	Q02470 lactobacill
5	261.5	6.5	1181	1 SCAZ_STRPY	P58099 streptococ
6	259	6.4	381	1 SUBT_BACST	P29142 bacillus st
7	257.5	6.4	1167	1 SCAL_STRPY	P15292 streptococ
8	257.5	6.4	1902	1 P3P_IACLC	P15292 lactococcus
9	257	6.4	381	1 SUBT_BACSU	P04189 bacillus su
10	257	6.4	1902	1 P1P_IACLC	P16271 lactococcus
11	255	6.3	381	1 SUBT_BACSA	P00783 bacillus su
12	254	6.3	381	1 SUBN_BACNA	P35835 bacillus su
13	252.5	6.3	1902	1 P2P_IACLC	P15293 lactococcus
14	248	6.2	380	1 ELYA_BACAO	P27693 bacillus al
15	248	6.2	380	1 ELYA_BACCS	P41362 bacillus cl
16	237.5	5.9	274	1 SUBD_BACLI	P00781 bacillus li
17	237.5	5.9	379	1 SUBT_BACLI	P00780 bacillus li
18	229	5.7	513	1 AQLI_THBAQ	P08594 thermus aqu
19	219.5	5.5	269	1 PRTM_BACSP	Q99405 bacillus sp
20	217	5.4	1398	1 PLS_PYRTU	P72186 pyrococcus
21	216.5	5.4	384	1 PRTK_TRIAL	P06873 tritirachiu
22	215.5	5.4	269	1 SUBT_BACLE	P29600 bacillus le
23	213.5	5.3	420	1 SUBT_BACSU	P28842 bacillus sp
24	213.5	5.3	1433	1 SUBF_BACSU	P16397 bacillus su
25	213	5.3	603	1 BPRV_BACNO	P42779 bacteroides
26	211	5.3	375	1 SUBT_BACPU	P07518 bacillus pu
27	211	5.3	319	1 ISPI_BACSU	P11018 bacillus su
28	209.5	5.2	269	1 SUBB_BACIE	P29339 bacillus le
29	206.5	5.1	388	1 CUDP_METAN	P29138 metathizium
30	205.5	5.1	395	1 BPRX_BACNO	P42780 bacteroides
31	204	5.1	378	1 ELYA_BACSP	P20724 bacillus sp
32	200	5.0	387	1 PRTR_TRIAL	P23653 tritirachiu
33	199.5	5.0	580	1 EXPR_XANCP	P23314 xanthomonas

34 199 5.0 1052 1 MS1P_HUMAN Q14703 homo sapien
35 198 4.9 1052 1 MS1P_CRIGR Q22248 cricetulus
36 196.5 4.9 467 1 ISP6_SCHPO Q40903 schizosacch
37 196.5 4.9 534 1 PROA_VIBAL P16588 vibrio algi
38 196 4.9 1052 1 MS1P_RAT Q9wt23 rattus norv
39 195 4.9 1052 1 MS1P_MOUSE Q9wt22 mus musculu
40 194 4.8 325 1 ISP_FAEPO P29139 paenibacill
41 193 4.8 422 1 TKSU_PYRKO P58502 pyrococcus
42 184 4.6 408 1 SEPR_THESR P80146 thermus sp.
43 183 4.6 533 1 PEPC_ASPNG P33295 aspergillus
44 181 4.5 409 1 ALP_TRIHA Q03420 trichoderma
45 178 4.4 401 1 THES_BACSP Q45670 bacillus sp

ALIGNMENTS

RESULT 1
SUBV_BACSU
ID SUBV_BACSU STANDARD; PRT; 806 AA.
AC P29141;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor extracellular protease vpr precursor (EC 3.4.21.-).
GN VPR OR IPA-45R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
RX MEDLINE=92041574; PubMed=1938992;
RA Sloma A., Ruto G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
RA Pero J.;
RT "Cloning and characterization of the gene for an additional
RT extracellular serine protease of Bacillus subtilis.";
RL J. Bacteriol. 173:6889-6895(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Tonescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruchli C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.F.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiesch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mollado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Roche M., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256 (1997).
CC -!- FUNCTION: NOT REQUIRED FOR GROWTH OR SPORULATION.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M76590; AAA22881.1; -;
DR EMBL; X73124; CAA51601.1; -;
DR EMBL; Z99123; CAB15835.1; -;
DR PIR; A41341; A41341;
DR HSP; P00782; ZSBI.
DR MEROPS; S08.00A; -;
DR Subtilisin; BG10591; vpr.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_s8.
DR Pfam; PF02225; PA; 1
DR Pfam; PF00082; Peptidase_s8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
DR Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 160
FT CHAIN 161 806 MINOR EXTRACELLULAR PROTEASE VPR.
FT ACT_SITE 189 189 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 534 534 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 806 AA; 85608 MW; F984E3BF0B969DDD CRG64;
Query Match 10.7%; Score 430; DB 1; Length 806;
Best Local Similarity 23.2%; Pred. No. 1.2e-21;
Matches 198; Conservative 135; Mismatches 299; Indels 222; Gaps 35;
QY 7 FLCIFLLFCSSEIL-----QKQYIVQLHPNS- 36
DB 8 FLLVSVFLPALSTGTVGQVQAPASSKTSADLEAEVFGDIDMTTSKTTIVLKEKSL 67
QY 37 -----ETAKTFASKFDWHLFLOEAVLGVVEEESPSRLLSYSGAIG 81
DB 68 AEAKEAGESQSKLKTARTKAK-----NKAIKAVK-----NGKNREYEQVFGS 112
QY 82 FAQLTSEAEILRYSEVVAVRP-----DHVLQVQTVSYKFLG--LDG-----FGNSG 129
DB 113 FSKMLPANELPKLAVKDKVAVPVNTVYKTDNKKDKDVTISDAVSPQMDGAPYIGAND 172
QY 130 VMSKSRFGQGTITGLVDITGVWPSPSPFDDTCMPSPRKKWGIQCEGESFSSSSCNKRLIG 189
DB 173 AWDLYGTGKIKVAIDITGVYHNPLD-----KKNFG 204
QY 190 ARFFIRGHRVANS---PEESP-NMPREYISARSTGHTASTVGGSSVSMANVLGNA 245
DB 205 Q---YKGYDFVNDYDKETPTGDP-----GEATDGHGTAVGVAAN----- 244
QY 246 GVARGMAPGAIHAYKVCWNGCVSSD-ILAAIDVAIQDKVDVLSLGGFFPIPLDYDTI 304

Db 245 GTIKVAPDATTLLAYRVLPGGSGTTENVIAGVERAVQDGMNLSGN---SLNNPDW 301
QY 305 AIGTFR--AMERGIVICAGNNGPIESSVAN--TAPWVSTIGAGTLDRRFPAV--RLA 358
Db 302 ATSTALDWNASEGVAVTSNGNSGPNGTWVSGPSTSRALISVGCATQLPLNEAVATFGSYS 361
QY 359 NGKLLYGESLYPGKIGKAGREVEVIYVTDGKGSEFCRLGSLPREIRGKMWICDRGVN 418
Db 362 SAKVMYGNKEDDVKALNN--KEVELV-EAGIGSEAKDF-----EGKDLTGKVAVVRGSI 412
QY 419 GRSEKGEAVKEAGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAVYNATVKPK 478
Db 413 AFVDKADNAKAGAGKGVVNNLSGIEANVPNGSVPTIKLSDEGEKL---VSALKAGE 469
QY 479 ARIIFGTVIGRGRAPDEVAQFSARGPSLANPSILKEDMIAPGVNIIAAMPQNLGPTGLPY 538
Db 470 TKITFKLTV--SKALGQVADFSRGP-VMDTWMIKPDISAPGVNIVSTIPTHDPDHPYGY 527
QY 539 DSRVNTVMSGTCMCPHVSGTALIRSAYPNWSPAATKSAUMTTA-DLYDRQGAIKD 597
Db 528 GSKQ-----GTSMASPHIAGAVAVIKQAPKWSVEQIKAAIMTAVTLKDSGDGEVYPH 590
QY 598 GNRPAGVFAIGAGHVPQKAINPLVYNIQPVVDYITVLTGLG-FTRSDILAIITHKN--- 652
Db 581 N-----AQGAGSARIMNAIKADSLVSPGYSYGTFLKENGNETKNETFTTENQSSIRK 633
QY 653 -----VSCNGILRKNPGFSL-----NYPSTAVIFKKGKT- 681
Db 634 SYLLEYSFNGSGISTGTSRVIPAHQTKATAKVKVNTKTKACTYEGTVIVREGKTV 693
QY 682 -----TEMITR-----RVTVN-----GSPNSIYSVNVKAPEG-----IKVIVNPKRL- 718
Db 694 AKVPTLLIVKEPDYPRVTSVSVSEGSVQGTYYQIETVLPAGAEELAFVLDVNSLDFAGQAG 753
QY 719 VFKAVDQTLRYRW 732
Db 754 IYKNQDKGYQYFDW 767
RESULT 2
SUBT BACAM
ID SUBT BACAM STANDARD; PRT; 382 AA.
AC P00782;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Subtilisin BPN precursor (EC 3.4.21.62) (Subtilisin Novo) (Alkaline
DE protease).
GN APR.
OS *Bacillus amyloliquefaciens*.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.
OX NCBI_TaxID=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23844;
EX MEDLINE=85006739; PubMed=6090391;
RA Vasantha N., Thompson L.D., Rhodes C., Banner C., Nagle J.,
RA Filpula D.;
RT "Genes for alkaline protease and neutral protease from *Bacillus*
RT *amyloliquefaciens* contain a large open reading frame between the
RT regions coding for signal sequence and mature protein.";
RL J. Bacteriol. 159:811-819 (1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84069812; PubMed=6316278;
RA Wells J.A., Ferrari E., Henner D.J., Estell D.A., Chen E.Y.;
RT "Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens*
RT subtilisin in *Bacillus subtilis*.";
RL Nucleic Acids Res. 11:7911-7925 (1983).
RN [3]
RP SEQUENCE OF 108-382.
RX MEDLINE=68086682; PubMed=6065094;
RA Markland F.S., Smith E.L.;

RT "Subtilisin BPN. VII. Isolation of cyanogen bromide peptides and the
 RL complete amino acid sequence.";
 RL J. Biol. Chem. 242:5198-5211(1967).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=72035041; PubMed=4399039;
 RA Alden R.A., Wright C.S., Kraut J.;
 RT "A hydrogen-bond network at the active site of subtilisin BPN.";
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:119-124(1970).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF COMPLEX WITH INHIBITOR.
 RX MEDLINE=85033707; PubMed=6387152;
 RA Hirono S., Akagawa H., Mitsui Y., Iitaka Y.;
 RT "Crystal structure at 2.6-A resolution of the complex of subtilisin
 BPN' with streptomyces subtilisin inhibitor.";
 RL J. Mol. Biol. 178:389-413(1984).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF MUTANT.
 RX MEDLINE=90057412; PubMed=2684274;
 RA Pantoliano M.W., Whitlow M., Wood J.F., Dodd S.W., Hardman K.D.,
 RA Rollence M.L., Bryan P.N.;
 RT "Large increases in general stability for subtilisin BPN' through
 RT incremental changes in the free energy of unfolding.";
 RL Biochemistry 28:7205-7213(1989).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX Gallagher T., Oliver J., Bott R., Betzel C., Gilliland G.L.;
 RT "Subtilisin BPN' at 1.6-A resolution: analysis for discrete disorder
 RT and comparison of crystal forms.";
 RL Acta Crystallogr. D 52:1125-1135(1996).
 RN [8]
 RP ACTIVE SITE.
 RX MEDLINE=69104413; PubMed=5249818;
 RA Markland F.S., Shaw E., Smith E.L.;
 RT "Identification of histidine 61 in the active site of subtilisin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 61:1440-1447(1968).
 CC -I- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in P1. Hydrolyzes peptide amides.
 CC -I- SUBCELLULAR LOCATION: Secreted.
 CC -I- BIOTECHNOLOGY: Used as a detergent protease. Sold under the name
 CC Alcalase by Novozymes.
 CC -I- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPORULATION, AND MANY MUTATIONS WHICH BLOCK SPORULATION AT EARLY
 CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPORULATION.
 CC -I- SIMILARITY: Belongs to peptidase family S8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; K02496; AAB05345.1; --
 DR PIR; B25415; SUBSN.
 DR PDB; 1S01; 15-OCT-90.
 DR PDB; 1S02; 15-JAN-92.
 DR PDB; 1SBH; 07-DEC-95.
 DR PDB; 1SBI; 07-DEC-95.
 DR PDB; 1SBN; 31-JAN-94.
 DR PDB; 1SBT; 31-MAY-84.
 DR PDB; 2SBT; 31-MAY-84.
 DR PDB; 1SIB; 31-OCT-93.
 DR PDB; 2SIC; 15-APR-93.
 DR PDB; 3SIC; 31-JAN-94.
 DR PDB; 5SIC; 31-JAN-94.
 DR PDB; 2SNT; 15-JAN-93.
 DR PDB; 1SPB; 15-OCT-95.

DR PDB; 2ST1; 15-JUL-91.
 DR PDB; 1ST2; 15-JUL-91.
 DR PDB; 1SUA; 14-JAN-98.
 DR PDB; 1SUB; 31-JAN-94.
 DR PDB; 1SUC; 31-JAN-94.
 DR PDB; 1SUD; 31-JAN-94.
 DR PDB; 1SUE; 14-OCT-98.
 DR PDB; 1SUP; 14-NOV-95.
 DR PDB; 1AK9; 12-NOV-97.
 DR PDB; 1AU9; 31-DEC-97.
 DR PDB; 1AQN; 14-JAN-98.
 DR PDB; 1YJA; 11-JUL-96.
 DR PDB; 1YJB; 11-JUL-96.
 DR PDB; 1YJC; 11-JUL-96.
 DR PDB; 1A2Q; 29-APR-98.
 DR PDB; 1DUI; 28-JAN-00.
 DR PDB; 1GNV; 24-JUN-02.
 DR PDB; 1LW6; 21-AUG-02.
 DR PDB; 1UBN; 28-JUN-99.
 DR MEROPS; S08.034; --
 DR InterPro; IPR000209; Peptidase S8.
 DR Pfam; PF00082; Peptidase S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal;
 KW 3D-structure.
 FT SIGNAL 1 32
 FT PROPEP 33 107
 FT CHAIN 108 382
 FT ACT_SITE 139 139
 FT ACT_SITE 171 171
 FT ACT_SITE 328 328
 FT HELIX 113 117
 FT TURN 118 119
 FT HELIX 120 126
 FT TURN 127 127
 FT TURN 131 132
 FT STRAND 134 139
 FT TURN 144 145
 FT TURN 147 148
 FT STRAND 151 156
 FT TURN 164 165
 FT HELIX 171 180
 FT TURN 193 194
 FT STRAND 196 201
 FT STRAND 203 203
 FT TURN 205 206
 FT STRAND 209 209
 FT HELIX 211 223
 FT TURN 224 225
 FT STRAND 228 231
 FT STRAND 235 235
 FT HELIX 240 251
 FT TURN 252 253
 FT STRAND 255 259
 FT STRAND 266 266
 FT TURN 267 268
 FT STRAND 269 269
 FT STRAND 274 274
 FT TURN 275 277
 FT TURN 279 280
 FT STRAND 282 287
 FT TURN 289 290
 FT STRAND 293 293
 FT TURN 295 296
 FT TURN 301 302
 FT STRAND 305 308
 FT STRAND 312 316
 FT TURN 317 319
 FT STRAND 320 324
 FT HELIX 327 344

```

FT TURN 346 347
PT HELIX 350 359
PT TURN 360 360
PT STRAND 362 362
PT HELIX 367 370
PT TURN 371 372
PT STRAND 374 374
PT HELIX 377 380
PT TURN 381 382
SQ SEQUENCE 382 AA; 39181 MW; ED987DAFA37B8335 CRC64;

Query Match 6.8%; Score 271.5; DB 1; Length 382;
Best Local Similarity 20.5%; Pred. No. 2.9e-11;
Matches 126; Conservative 63; Mismatches 177; Indels 249; Gaps 17;

QY 6 FFLCIIFLFCSSSEILQKQYIVOLHPNSTATKAFKPDWHLSTFQBAVLGVBESEE 65
DB 12 FALALIFTMAFGSTSS-----AQAGKSGEKYIVGFQKTWTSTMSAAKKDKDVISEK 63

QY 66 EBSRLLYSGSAIEGFAQLTESEAEILRYPEVAVRDPDHLVQVQTTYSYKF-LGLDGG 124
DB 64 GKGKQKQFYK---VDAASATLNEKAVKELKDPSPVAYVEEDHVAH---AYAQSVDPYGVSQ 117

QY 125 FGNSGVWSKSRFGQGTIIIGVLDTGVMPSPSPFDDTGMPSPKWKGIQCEGESFSSSCN 184
DB 118 IKAPALHSQGYTGSNVKVAIVDSGIDSSHPDLKVAGGASM----- 157

QY 185 RKLIGARFTIRHRVANSPEESPMPREYISARDSTGHGTHASTVGGSSVSMANVLNG 244
DB 158 -----VPSETNPQDNNSHGTHVAGTVA-----ALNNS 185

QY 245 AGVARGMAPGAIHATYKVCWFNGC-YSSDILAAIDVAIQDKVDVLSLGGPFPLYDDT 303
DB 186 IGVL-GVAPSASLYAVKVLGADGGQGYWIINGIEWAIANNMDVINMSLGG---PSGSAA 241

QY 304 IAIQTFRAMEGTSVICAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLANGKLL 363
DB 242 LKAANDKAVASGVVVVAAGNETSGSS-----STVG----- 273

QY 364 YGSESLYPGKGIKNAGREVEIYVTGGDKGSEFCLGSLPREIRGMVICDRGVNGRSEK 423
DB 274 -----YPGK----- 277

QY 424 GEAVKGAGGVAMILANTEINQBEDSDVHLHPATL-IGYTESVLLKAYNAVTKPKARII 482
DB 278 -----YPSVIAVGAVDS----- 289

QY 483 FGGTVIGRSEAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGLPYDSRR 542
DB 290 -----SNQRAGFSVYGPGL-----DVMAPGVSIQSTLPGN----- 319

QY 543 VNFVMSGTSWSPHYSGITALIRSAYPNWSAAIKSALMTTADLYDRQKAIKDNKPA 602
DB 320 -KYGAYNGTSMASHVAGAAALILSKHPNWTNTQVRSSLENTT-----TKLG 365

QY 603 GVFAIGAGHVNPKQA 617
DB 366 DSFYVKGKGLNVQAA 380

RESULT 3
ID ELYA BACHD STANDARD; PRT; 361 AA.
AC P41363; C53294; Q9KEJ7;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
GN Thermostable alkaline protease precursor (EC 3.4.21.-).
DN BH0855.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=85665;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=AH-101 / JCM 9161;
RX MEDLINE=9309826; PubMed=1369007;
RA Takami H., Kobayashi T., Aono R., Horikoshi K.;
RT "Molecular cloning, nucleotide sequence and expression of the
RL structural gene for a thermostable alkaline protease from Bacillus
RN sp. no. AH-101.";
RP Appl. Microbiol. Biotechnol. 38:101-108(1992).
RP [2]
RP REVISIONS.
RA Takami H., Kobayashi T., Aono R., Horikoshi K.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RL Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RN Horikoshi K.;
RP "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
RN [4]
RP SEQUENCE OF 94-113.
RC STRAIN=AH-101 / JCM 9161;
RX MEDLINE=91103969; PubMed=1370008;
RA Takami H., Akiba T., Horikoshi K.;
RT "Characterization of an alkaline protease from Bacillus sp. no.
RL AH-101.";
RN Appl. Microbiol. Biotechnol. 33:519-523(1990).
CC -1- FUNCTION: SHOWS KERATINOLYTIC ACTIVITY AND HAS A HIGH OPTIMUM PH
(CC (PH 12-13)).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13158; BAA02443.2; -.
CC EMBL; AP001510; BAB04574.1; -.
CC F1; G83756; G83756.
CC HSSP; Q99405; 1MPT.
CC MEROPS; S08.00A; -.
CC InterPro; IPR00209; Peptidase S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE ASP; 1.
CC PROSITE; PS00137; SUBTILASE HIS; 1.
CC PROSITE; PS00138; SUBTILASE SER; 1.
CC Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
KW SIGNAL 1 24
FT PROPEP 25 93
FT CHAIN 94 361
FT ACT_SITE 124 124
FT ACT_SITE 154 154
FT ACT_SITE 307 307
FT VARIANT 38 38
FT VARIANT 69 70
FT VARIANT 74 77
FT VARIANT 86 86
FT VARIANT 104 104
FT VARIANT 292 292
FT VARIANT 346 347
SQ SEQUENCE 361 AA; 38116 MW; 24BF004F9E3E8474 CRC64;

Query Match 6.8%; Score 264.5; DB 1; Length 361;
Best Local Similarity 19.6%; Pred. No. 8.2e-11;
Matches 115; Conservative 73; Mismatches 141; Indels 259; Gaps 18;

```



```

QY 295 FP--IPLYDDTTAIGTFRAMEGSGISVICAAGNNGPIESS-----VAN 334
Db 353 DSGNQYLEDPEIA-AVONANESGTAIVISAGNSGTSATQGVNDKYVGLQDNEMVGP 411
QY 335 TAPWSTIGAGTLDRFPVAVRLANGKLYGSLYPGKIGKAGREVEIYVYTGDKGSE 394
Db 412 TSGGATVSAENTDVISQAVITDKDL--QLGP-----ETIQISSLNDFGTS 457
QY 395 FCUR-----GSLPR-----BEIRGKVICDRGVNGRSEKGEAVKAGGVAMILA 438
Db 458 FDKQKPYVVKDAGSLGAAADYTADAKGIAIVKRGELNADFQKVAQAAGAAGLIIV 517
QY 439 NTEINEEDSIDVHLPLATLIGTVESV-----LLKAYVNAVTKPKARIIFGTV 487
Db 518 NN-----DGTATPLSIRLITTPFFGLSSGTQKLVDTVAHDDSLGVKIALTL 568
QY 488 IGRSRAPE--VAQFSARGPSLANPSILKPDMLAPGVNIIAANPQLGTLFPYDSRRVNF 545
Db 569 LPNQKTEDKMSDFTSYGP-VSNLS-FKPDITAPGNIWSTQNN-----GY 613
QY 546 TVMSGTSMCPHVSIGTALIRGAYPNWSPAAIKSALMTTADLYDRQKAIKDNKPKGVF 605
Db 614 TMSGTSMASPPFTAGSQALLQALNNKNPFY-----ADYKQLGTALTDFLKTVE-- 664
QY 606 AIGAGHVPQKAINPGLVNIQVDYITLCTLGTRSDILAIHKNVSCGILKPNPF 665
Db 665 -----MNTAQINDINNNVIVSPRQAGLVKDVKAIDA-LEKNPST 706
QY 666 SL---NYPPIAV-----IPK---RGKTTEMITRRVTVNGSPNSIYVNVKADEGI 709
Db 707 VVAENGYPAVELEKDFSTDKTKFLFTNTHLYQVDSNTDTHAVTISATDPNSGV 764

```

RESULT 5

```

SCA2_STRPY
ID SCA2_STRPY STANDARD; PRT: 1181 AA.
AC PS8099;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CSA peptidase precursor (EC 3.4.21.-) (SCP).
GN SCFA OR SPY2010.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Szate S., Suvarov A.N., Keston S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus
RT pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -!- FUNCTION. THIS VIRULENCE FACTOR OF S. PYOGENES SPECIFICALLY CLEAVES
CC THE HUMAN SERUM CHEMOTAXIN CSA AT LYS(68)-ASP(69) BOND NEAR ITS
CC C-TERMINUS, DESTROYING ITS ABILITY TO SERVE AS A CHEMOATTRACTANT.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL; AE006623; AAK34691.1; -.
DR HSP; Q45670; IDBI.
DR MEROPS; S08.020; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXIG.
DR InterPro; IPR003137; PA.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; FALSE NEG.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Repeat;
KW Signal; Complete proteome.
FT SIGNAL 1 31 BY SIMILARITY.
FT CHAIN 32 1147 CSA PEPTIDASE.
FT PROPEP 1148 1181 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 130 130 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 512 512 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1029 1121 CELL WALL ATTACHMENT DOMAIN (POTENTIAL).
FT DOMAIN 1034 1118 5 X 17 AA TANDEM REPEATS.
FT REPEAT 1034 1050 1.
FT REPEAT 1051 1067 2.
FT REPEAT 1068 1084 3.
FT REPEAT 1085 1101 4.
FT REPEAT 1102 1118 5.
FT SITE 1144 1148 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1147 1147 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1181 AA; 129486 MW; 39FC51763419CFC CRC64;

Query Match 6.5%; Score 261.5; DB 1; Length 1181;
Best Local Similarity 24.2%; Pred. No. 6; Se-10;
Matches 128; Conservative 95; Mismatches 198; Indels 107; Gaps 26;

QY 133 KSRFGQGTIGVLDTGWPSPSFDTGMP-SIPRKWKGICQEGESFSSSSCNKRLIGAR 191
Db 117 KAGKGAGTVVAVIDAG-----FDKNHEAWRLTDKTKARYQSKEDLEKA---KKEHGIT 166
QY 192 FFRGHRVANSPEESPMPREVISARDSTGCHGHTASTVGGSSVSMANVLGNGAGVARGM 251
Db 167 Y---GEWVDKVAHYHDYSKDGKTAVDQE-HGHVSGILSGNAPSETKBPYR-----LEGA 218
QY 252 APGAHIAVYKVCWFNGC--YSSDILAAIDVAIQKVDVLSLSLGGFPPIP---LYDDTIAI 306
Db 219 MPEAQILLMEVEIVNGGLADYARNYAQAIDAVNLGAKVINMSFGNAALAYANLPDETKKA 278
QY 307 GTFRAMEGISTVCAAGNNG-----PI-----ESSVANT---APWVSTIGAGTLDRFP 351
Db 279 FDY-AXSKGVSIVTSAGNDSFGGKTRPLADHPDYGWVGTPAAADSTLTVASYSPPDKQL 337
QY 352 --PAVVELANGK-----LLYGESLYPGKIGKAGREVEIYVYTGDKGSEFCRLGSLPRE 404
Db 338 TETATVXTAQDQDKEMPVLSTNFEPNKAYDYA-----YANRGMKEDDF-----K 382
QY 405 EIRGKVICDRGVNGRSEKGEAVKAGGVAMILANTEINQEE-----DSIDVHLPLATLIG 460
Db 383 DVKGKIALIERGDIDFKDKIANAKKAGAVGLIYD---NODKGFPIELPNVDQMPAFIS 439
QY 461 YTESVLLKAYVNATVKPKARIIFGGT--VIGRSRAPEVAQFSARGPSLANPSILKPDMLIA 518
Db 440 RKDGLLKE-----NPQKITITENATPKVLPASGTKLSRFSSWG--LTADGNKIKPDIAA 491
QY 519 PGVNIIAANPQNLGPTGLPYDSRRVNTVMGTSMSCPHVSIGTALIR-----SAYNWSNP 574
Db 492 PGQDILSSVANN-----KYAKLSGTSMSAPLVAIGLGLQKQYETQYEDMTP 538
QY 575 A----AIKSALMTAD-LYDRQKAIKDNKPKGVFAIGAGHVPQKA 617
Db 539 SERLDLAKKVLMSATALYDEDEKAFSPRQK-----GAGAVDAKKA 580

```

RESULT 6

SUBT_BACST STANDARD; PRT; 381 AA.

AC P29142; 21-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Subtilisin J precursor (EC 3.4.21.62).

GN APRJ

OS Bacillus stearothermophilus.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.

OX NCBI_TaxID=1422;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIMB 10278 / KCTC 1823;

RX MEDLINE=92231938; PubMed=1567435;

RA Jang J.S., Kang D.O., Chun M.J., Byun S.M.;

RT "Molecular cloning of a subtilisin J gene from Bacillus

RL Biochem Biophys Res Commun. 184:277-282(1992).

CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,

CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity

CC for peptide bonds, and a preference for a large uncharged residue

CC in pl. Hydrolyzes peptide amides.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF

CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY

CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN

CC IS NOT NECESSARY FOR NORMAL SPOULATION.

CC -!- SIMILARITY: Belongs to peptidase family S8.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M64743; AAA22247.1; -

DR PIR; J01487; J01487.

DR HSP; P04189; ISCU.

DR MEROPS; S08.035; -

DR InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE_ASP; 1.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolyase; Sporulation; Serine protease; Zymogen; Signal.

FT SIGNAL 1 29 POTENTIAL.

FT PROPEP 30 106 POTENTIAL.

FT CHAIN 107 381 SUBTILISIN J.

FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).

SQ SEQUENCE 381 AA; 39495 MW; C7A596F7629087D5 CRC64;

Query Match 6.4%; Score 259; DB 1; Length 381;

Best Local Similarity 21.2%; Pred. No. 2.1e-10;

Matches 130; Conservative 61; Mismatches 175; Indels 248; Gaps 20;

QY 6 FELCIPLFCSSSEILOQTIVVOLHPNSETAKTFASKDWHLSFLQEAFLGVETEE 65

DB 12 FAULTPLTFMFSNWS-----VQAAGKSTTEKKYIVGFKQTMASAKKDVISBK 62

QY 66 EPSRLLYSYGSAIEGFAALTESEAEILRYSPEVAVRPDHLVQLQVTTYSYKFLGLDGF 125

DB 63 GGRVQKQPKY---VNAATLDEKAVKELKQDSVAYVEEDHIAH-EYAQSVPY-GISQI 117

QY 126 GNSGVWSKSRFGQGTIIIGVLDTCVWPSPSFDGTGMPSPRKKWKIGICQEGESFSSSCNR 185

DB 118 KAPALHSQGYTGSNVKVAVIDSGIDSSHPLNVRG----- 152

QY 186 KLJGAREFFIRGHRVANSPEESPNNMREYISARDSTGHTGTHASTVGGSSVSMANVLNGA 245

DB 153 ---GASP-----VPSETNDYQDSSHGTHVAGTIA-----ALNNSI 185

QY 246 GVARGMAPGAHIAVYKVCWFGNC-YSSDILAAIDVAIQKVDVLSLSLGGFFIPLYDDTI 304

DB 186 GVL-GVSPSASLVAVKLDSTGSGQYSWIINGIEWALSNNMVDVNNLSLG---PSGSTAL 241

QY 305 AIGTFRAMEGIVSICAGNNGPIESSVANTAPWVSTIGAGTLDLRPPAVVRLANGKLLY 364

DB 242 KTVVDKAVSSGIVVAAAAGNEGSGS-----STVG----- 272

QY 365 GESLYPGKGIKNAGREVEVYVGGDKGSEFCGLSPREEIRGMKVICDRGVNGRSEKG 424

DB 273 ---YPAK----- 276

QY 425 EAVKEAGGVAMILANTEINQEDSIDVHLLPATL-IGYTESVLLKAYVNAVTKPKARIIF 483

DB 277 -----YPSIIAVGAVNS----- 288

QY 484 GGTVIGSRAPVAQSFARGPSLANPILKPDIMAFGVNIIAAMPQNLGPTGLPYDSRV 543

DB 289 -----SNQASFSAGSEL-----DWAPGVSI-----QSTLPGG----- 318

QY 544 NPTVMSGTSCHVSGITALIRSAVPNWSPPAKKALMTADLYDRQGAIKDGNKPGAG 603

DB 319 TYGAYNGTSNATPHVAGAAALILSKHPTWTNAQVRDLRESTATYL-----GNS--- 366

QY 604 VFAIGAGHVNPQKA 617

DB 367 -FYVKGGLINQAA 379

RESULT 7

SCAL_STRPY STANDARD; PRT; 1167 AA.

AC P15926;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE CSA peptidase precursor (EC 3.4.21.-) (SCP).

GN SCPA.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

CC Streptococcus.

OX NCBI_TaxID=1314;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-41.

RX MEDLINE=90153964; PubMed=2406246;

RA Chen C.C., Cleary P.P.;

RT "Complete nucleotide sequence of the streptococcal C5a peptidase gene

RL J. Biol. Chem. 265:3161-3167 (1990).

CC -!- FUNCTION: THIS VIRULENCE FACTOR OF S. PYOGENES SPECIFICALLY CLEAVES

CC THE HUMAN SERUM CHEMOTAXIN C5A AT LYS(68)-ASP(69) BOND NEAR ITS

CC C-TERMINUS, DESTROYING ITS ABILITY TO SERVE AS A CHEMOATTRACTANT.

CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by

CC an amide bond (potential).

CC -!- SIMILARITY: Belongs to peptidase family S8.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; J05229; AAA26960.1; -


```

DR PIR; A35066; A35066.
DR HSP; Q45670; 1DBI.
DR MEROPS; S08.020; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006132; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Repeat;
KW Signal.
FT SIGNAL 1 31
FT CHAIN 32 1130 CSA PEPTIDASE.
FT PROPEP 1131 1167 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 130 130 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 512 512 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1029 1104 CELL WALL ATTACHMENT DOMAIN (POTENTIAL).
FT REPEAT 1034 1050 4 X 17 AA TANDEM REPEATS.
FT REPEAT 1051 1067 1.
FT REPEAT 1068 1084 2.
FT REPEAT 1085 1101 3.
FT SITE 1127 1131 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1130 1130 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT SEQUENCE 1167 AA; 128263 MW; D2DDC5E5752DASD CRC64;

Query Match 6.4%; Score 257.5; DB 1; Length 1167;
Best Local Similarity 24.1%; Pred. No. 1.2e-09;
Matches 127; Conservative 96; Mismatches 198; Indels 107; Gaps 26;

QY 133 KSRFGQGTIGVLTGWVPSFDDTQMP-SIPRKWKICQEGESFSSSCNRKLIGAR 191
DB 117 KAGKAGTVAVIDAG-----FDKNHEARLTDKTKARYQSKDELEKA---KKEHGTT 166
QY 192 FPIRGHVRVANSPEESPMNPREYTSARDSTGHGHTASTVGGSSVANVLGNAGVARGM 251
DB 167 Y---GEWNDKVAYHYDYSKGTAVDOE-HGTEVSGILSGNAPSTKPYR-----LEGA 218
QY 252 APGAHIAVYKVFENGCC--YSSDILAAIDVAIQKVDVLSLGGPPIP---LYDDTIAI 306
DB 219 MPEAQLLMREVEIVNGLADYARNVAQAIRDAVNLGAKVINMSPGNAALAVANLPDETKA 278
QY 307 GTFRAMERGISVICAANG-----PI-----ESSVANT---APVSTIGAGTLDRFP 351
DB 279 FDY-AKSKGVSIIVTSAGNDSFSGKTRPLADHPDYGVVGTAAADSTLTVASYPDKQL 337
QY 352 --PAVVLANKG-----LLYGESLYPGKGIKNAGREVEIYVYTGDKGSEFCLRGLPRE 404
DB 338 TETAMVATDQDQKEMPEVLTNPFEPNKAIDYA-----YANRGMKEDDF-----K 382
QY 405 EIRGKMWICDRGVNGRSEKGEAVKEAGGVAMILANTEINQEE----DSIDVHLPLATLIG 460
DB 383 DVKGKIALIERGDDIDFKVANAKKAGAGVGLYD---NODKGFPIELFNVDPMPAAFTS 439
QY 461 YTESVLLKAYNATVTKPARIIFGGT--VIGRSRAPEVAQPSARGPSLANPSILKPDMTA 518
DB 440 RKDGLLLKD-----NPKQTITFNATPKVLPTASGTLKSRFSWG--LTADGNKPKDIAA 491
QY 519 PGVNIIIAWPNQLGPTGLPYDSRVNFTVMSGTSMSCPHVSIGTILIR-----SAYPNWSP 574
DB 492 PGQDILSSVANN-----KYAKLSGTSMSAPLVAGIMGLLQKQYETQYPDWTF 538
QY 575 A----ALKSALMTTAD--LYDRQGGKAIKDGKNKPGVFAIGAGHVNPOKA 617
DB 539 SERLDLAKVLVMSSTALYDEDEKAYFSPRQQ-----GAGAVDAKKA 580

```

```

RESULT 8
P3P_LALCLC
ID P3P_LALCLC STANDARD; PRT; 1902 AA.
AC P15292; 1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PIII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
associated serine proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Plasmid.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK11;
RA MEDLINE=89340435; PubMed=2760036;
RX Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:13579-13585(1989).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
specificity, although some substrate preference have been noted,
e.g. large hydrophobic residues in the P1 and P4 positions, and
Pro in the P2 position. Best known for its action on caseins, and
although it has been shown to hydrolyze hemoglobin and oxidized
insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
an amide bond (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J04962; AAA03533.1; ALT_SEQ.
CC HSPF; P00782; 2SET.
CC MEROPS; S08.019; -.
CC InterPro; IPR001899; Gram_pos_anchor.
CC InterPro; IPR006192; LPXTG.
CC InterPro; IPR003137; PA.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00746; Gram_pos_anchor; 1.
CC Pfam; PF02225; PA; 1.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
CC PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
CC PROSITE; PS00136; SUBTILASE_ASP; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW Signal; Plasmid.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1870
FT PROPEP 1871 1902 PIII-TYPE PROTEINASE.
FT ACT_SITE 217 217 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 1867 1871 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1870 1870 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;

Query Match 6.4%; Score 257.5; DB 1; Length 1902;

```

Best Local Similarity 21.08; Pred. No. 2.3e-09;
Matches 162; Conservative 124; Mismatches 287; Indels 197; Gaps 38;

```
QY      17 SSSEIIKQKTYIVQLHPNSETAKTASFKDWHLSFLQEAVLGVEEESPESSRLLYSYG 76  
DB      115 SSTAETIQ-----ETNKVIRAQ-----ASVKAAVEQTQQTAGE---SVG 151  
  
QY      77 SAIEGFAAQLTSEAEILRYSPEVAVRPDHVLQVOTYSKYFLGLDGFGNS-----GVW 131  
DB      152 YVNGFSTKVRVDIPKLK--QIAGVK-----TVTLAKVYYPTDAKANSMANVOAWM 201  
  
QY      132 SKSRP-CQGTHIGVLDGVWPESPSP---DDTGM---PSIPRWKGICQEGSFSSSCN 184  
DB      202 SNYKYKEGWVSVIDSGIDPTHKMRSLSDDKOVKLTSDVEKFDTVTKHGRFYNSK---258  
  
QY      185 RKLIGARFFIRGHRVANSPESPMPREYTSARDSTGHGHTASTVTGGSSVSMAVLGNL 244  
DB      259 ---VPYGF-----NYADNNITTDKVEDQ-----HGHWAGIIGAN-----GTG 295  
  
QY      245 AGVAR---GMAPGAHIAVYKCVFNG-----CYSSDILAIDAIVAIQDKVDLSLGLGF 295  
DB      296 DDEPAKS VVGVAPEAQLAMKV---FSNSDTS AKTGSATVVSIAEDSAKIGADVLMNLSEN 353  
  
QY      296 P-IPLYDDTIAGTFRAMERGISVICAAGNCGPIESSVANTAPVY-----STIAGTFL 347  
DB      354 SGNQTLDEPELA-AVQNANESGTAAVISAGNSGTSATGEVKNKYGLQDNEMWGSPGT 412  
  
QY      348 DRFPFVAVRLANGKLL-VGESLVPGKIGKNAGREVV-----IYVTGGDKGS 393  
DB      413 SRGATTVASRENTDVITQAVITDGTGLQGPETIQLSHDFTCGSPDQKFIIVXDASN 472  
  
QY      394 EFLCROSLP---REEIRGOWICDRGVNGRSEKGEAVEKEAGGVAMILANTEINQEEIDSIV 451  
DB      473 --LSKALADYTADAKGKIAVKRGFEPSFDKKYAQAAGAAGLIIVNT-----DG 521  
  
QY      452 HLPATLIGVT-----ESULLKAYVN-ATVKPARIFGTVIGRAP-----E 495  
DB      522 TATEMTSIALTTFTPTFGLSSVVGQKLVDMVTAHPDDSL---GVKITLAMLNQKYTEDK 578  
  
QY      496 VAQFSARGPLANSFILKPDMAIPGVNIATAAPQNLGTPGLPYDSRRVNFVWSGTSMSC 555  
DB      579 MSDFTSYGP-VSNLS-FKPDITAPGNIWSTQNN-----GYTNWSGTSMSAS 623  
  
QY      556 PHYSGITALIRSYPNPNSPAIAKLSALMTTADLYDRQ--GKAIKDGNKPAGVFAIGAGHYN 613  
DB      624 PFTAGSQALKQALNNKN-----NPFYAYYKQKGTALTDLFKTYE-----664  
  
QY      614 PQKAINPGLVYNTPVDYITYLTCLGFTSRDILAITHKNVCSNGILRKNPGPSL---NVP 670  
DB      665 -----MNTAQPINDIANNVNVISPRRQAGLVDMKAAIDA-LEKNPSTVWAENGYP 714  
  
QY      671 SIAYV-----LFK---RGKTTEMITRRVTVNVSFNISYVNVXAPEGI 709  
DB      715 AVELKDFTSTDKTFKLTFTNRTHIELTYQDNSNTDNAVYTSATDPSNSGV 764
```

RESULT 9

ID	SUBT_BACSU	STANDARD;	PRT;	381 AA.
AC	P04189; C007613; P70989;			
DC	20-MAR-1987 (Rel. 04, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Subtilisin E precursor (EC 3.4.21.62).			
GN	APRE OR APR OR APRA OR SPRE.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.			
NCBI_TaxID=1423;				
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RC	MEDLINE=84212198; PubMed=6427178;			
RX	"The complete genome sequence of the Gram-positive bacterium Bacillus subtilis."			
RT	J. Bacteriol. 171:2657-2665(1989).			
STahl M.L., Ferrari E.;				

"Replacement of the Bacillus subtilis subtilisin structural gene with an in vitro-derived deletion mutation."
J. Bacteriol. 158:411-418(1984).
[2]
SEQUENCE FROM N.A.
STRAIN=168;
MEDLINE=98240224; PubMed=9579061;
Noback M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R., Wedler H., Venema G., Bron S.;
"The 172 kb prkA-addAB region from 83 degrees to 97 degrees of the Bacillus subtilis chromosome contains several dysfunctional genes, the glyB marker, many genes encoding transporter proteins, and the ubiquitous hit gene."
Microbiology 144:859-875(1998).
[3]
SEQUENCE FROM N.A.
STRAIN=168;
MEDLINE=98044033; PubMed=9384377;
Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriss L., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Coffeau A., Golightly E.J., Grandi G., Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blandford M., Klein C., Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.N., Levine A., Liu H., Masuda S., Maue C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieser M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vanderbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
"The complete genome sequence of the Gram-positive bacterium Bacillus subtilis."
Nature 390:249-256(1997).
[4]
SEQUENCE OF 1-156 FROM N.A.
STRAIN=168 / PY79; PubMed=632190;
MEDLINE=84144862; PubMed=3108260;
Wong S.L., Price C.W., Goldfarb D.S., Doi R.H.;
"The subtilisin E gene of Bacillus subtilis is transcribed from a sigma 37 promoter in vivo."
Proc. Natl. Acad. Sci. U.S.A. 81:1184-1188(1984).
[5]
SEQUENCE OF 1-156 FROM N.A.
STRAIN=168 / PY79;
MEDLINE=87222417; PubMed=3108260;
Ikemura H., Takagi H., Inouye M.;
"Requirement of pro-sequence for the production of active subtilisin E in Escherichia coli."
J. Biol. Chem. 262:7859-7864(1987).
[6]
SEQUENCE OF 1-13 FROM N.A.
STRAIN=168;
MEDLINE=85213955; PubMed=2496113;
Park S.-S., Wong S.L., Wang L.F., Doi R.H.;
"Bacillus subtilis subtilisin gene (aprE) is expressed from a sigma A (sigma 43) promoter in vitro and in vivo."
J. Bacteriol. 171:2657-2665(1989).

RN SEQUENCE OF 1-8 FROM N.A.
 RP MEDLINE=88068685; PubMed=2447063;
 RX Hemmer D.J.; Ferrarri E.; Perego M.; Hoch J.A.;
 RA "location of the targets of the hpr-97, sacJ32(Hy), and sacQ36(Hy)
 RT mutations in upstream regions of the subtilisin promoter."; J.
 RL J. Bacteriol. 170:296-300(1988).
 [8]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RC STRAIN=168;
 RX MEDLINE=99030466; PubMed=9811547;
 RA Jain S.C.; Shinde U.; Li Y.; Inouye M.; Berman H.M.;
 RT "The crystal structure of an autoprocessed Ser221Cys-subtilisin
 RT E-propeptide complex at 2.0-A resolution."; J.
 RL J. Mol. Biol. 284:137-144(1998).
 CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in fl. Hydrolyzes peptide amides.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPORELLATION, AND MANY MUTATIONS WHICH BLOCK SPORELLATION AT EARLY
 CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPORELLATION.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; K01988; AAA22742.1; -
 CC ENBL; Y14083; CAA74536.1; -
 CC ENBL; Z99109; CAA12870.1; -
 CC ENBL; K01443; AAA22814.1; -
 CC ENBL; M16639; AAA22744.1; -
 CC ENBL; M31060; AAA22246.1; -
 CC ENBL; M19125; AAA22245.1; -
 CC PIR; A00972; SUBSI.
 CC PDB; 1SCU; 13-JAN-99.
 CC MEROPS; S08.036; -
 CC Subtilist; BG10190; abpr.
 CC InterPro; IPR000209; Peptidase_S8.
 CC Pfam; PF00082; Peptidase_S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC PROSITE; PS00136; SUBTILASE ASP; 1.
 CC PROSITE; PS00137; SUBTILASE HIS; 1.
 CC PROSITE; PS00138; SUBTILASE SER; 1.
 CC Hydrolase; Sporulation; Serine protease; Zymogen; signal;
 KW 3D-structure; Complete proteome.
 KW SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 106 POTENTIAL.
 FT CHAIN 107 381 SUBTILISIN E.
 FT ACT_SITE 138 138 CHARGE RELAY SYSTEM.
 FT ACT_SITE 170 170 CHARGE RELAY SYSTEM.
 FT ACT_SITE 327 327 CHARGE RELAY SYSTEM.
 FT CONFLICT 27 27 V -> A (IN REF. 1, 4 AND 5).
 FT STRAND 37 43
 FT STRAND 52 60
 FT HELIX 61 63
 FT TURN 65 69
 FT STRAND 75 80
 FT STRAND 82 89
 FT TURN 90 90
 FT TURN 92 93
 FT TURN 94 99
 FT STRAND 102 105
 FT STRAND 112 116
 FT HELIX 117 118

FT HELIX 119 125
 FT TURN 126 131
 FT TURN 133 138
 FT STRAND 143 144
 FT TURN 146 147
 FT STRAND 150 155
 FT TURN 158 159
 FT TURN 163 164
 FT TURN 170 179
 FT TURN 192 193
 FT STRAND 195 200
 FT TURN 204 205
 FT STRAND 207 209
 FT STRAND 210 222
 FT TURN 223 224
 FT STRAND 227 230
 FT STRAND 232 234
 FT HELIX 239 250
 FT TURN 251 252
 FT STRAND 254 258
 FT STRAND 265 265
 FT TURN 266 267
 FT STRAND 268 268
 FT STRAND 273 273
 FT TURN 274 276
 FT TURN 278 279
 FT STRAND 281 286
 FT TURN 288 289
 FT STRAND 292 292
 FT TURN 294 295
 FT TURN 300 301
 FT STRAND 304 307
 FT STRAND 311 315
 FT TURN 316 318
 FT STRAND 319 323
 FT HELIX 326 343
 FT TURN 345 346
 FT HELIX 349 358
 Query Match 6.4%; Score 257; DB 1; Length 381;
 Best Local Similarity 21.0%; Pred. No. 2.9e-10;
 Matches 129; Conservative 62; Mismatches 175; Indels 248; Gaps 20;
 QY 6 PFLCIIFLLFCSSSSSEILQQTIVVQLHPNSETAKTFASKFDWHLSPLOEAVLGVSEEE 65
 DB 12 FALTLIIFTMAFSNMS-----VQAAGKSTSEKKYIVGFKQTMSSAMSAKKKDVISEK 62
 QY 66 EPSRRLLYSYGSAIEGFAAQLTSEAEILRYSPEWAVRDPHVLRQVTTYSYKFLGIDGF 125
 DB 63 GGRVKQKQFKY---VNAAAATLDEKAVKELKDFSAVVEEDHIAH-EYASQVY-GISQI 117
 QY 126 GNSGVMSKSRPQGQTIIGVLDTGWVPSPSPFDDTGMPSIPRKWKGICQEGESFSSSCNR 185
 DB 118 KAPALHSQGTGSNVKVAVIDSGIDSSHPLNVRG----- 152
 QY 186 KLICARFFIRGHRVANSPESSPNMPREYISARSTGHGTATSTVGSSSVSMANVLNGA 245
 DB 153 ---GASF-----VPSETNPYQDGSSHGTHVAGTIA-----ALNNSI 185
 QY 246 GVAEGMAPGAHIAVYKVCWFNGC-YSSDILLAAIDVAIQDKVDVLSLGLGFPPIPLYDDTI 304
 DB 186 GVL-GVSPFASLYAVKVLDDTSGSQYSWIINGIEWAISSNNMDVINMSLGG---PTGSTAL 241
 QY 305 AIGTFRAMEGIVSICAAAGNNGPIESSVANTAPWVSTIGAGTLDLRRFPVAVRLANGKLLY 364
 DB 242 KTVVDKAVSSGIVVAAAGNEGSSGST-----STVG----- 272
 QY 365 GESLYPGKIKNAGREVEVIYVTGDKGSEFCLRGSLPREIRGNVICDRGVNGRSEKG 424
 DB 273 ---YPAK----- 276
 QY 425 EAVKEAGGVAMILANTEINQEEDSIDVHLLPATL-IGYTESVLLKAYVNAVTKPKARIIF 483

```

Db 277 -----YPSIIAGVANS----- 288
Qy 484 GGTVIGRSRAPEVAQFARGSPSLANPSILKPDMPAGVNIIAAWPQNLGPTGLPYDSRRV 543
Db 289 -----SNQRAFSSAGSEL-----DVAAPGVS-----QSTLPGG----- 318
Qy 544 NTFVMSGTSMSCHVSGITALLIRSVNWSPAKAKALMTADLYDRQKAIKDGKNKAG 603
Db 319 TYGAYNGTSMATPHVAGAAALILSKIPTWNAQVRDLSESTATYL-----GNS----- 366
Qy 604 VFAIGAGHVNPOKA 617
Db 367 -FYGKGLINVOQA 379

RESULT 10
PIP_LACLC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P1-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
DE proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pW05.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WG2;
RA MEDLINE=98149035; PubMed=3278687;
RA Kok J., Leenhouck K.J., Haandrikman A.J., Ledebor A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
RT Streptococcus cremoris Wg2.";
RL Appl. Environ. Microbiol. 54:231-238(1988).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC -!- GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC Pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyze hemoglobin and oxidized
CC insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24767; AAA17677.1; --
CC DR HSSP; P00782; IS01.
CC DR MEROPS; S08.019; --
CC InterPro; IPR001899; Gram_pos_anchor.
CC InterPro; IPR006192; LPXTG.
CC DR InterPro; IPR003137; PA.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00746; Gram_pos_anchor_1.
CC Pfam; PF02225; PA; 1.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
CC PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
CC PROSITE; PS00136; SUBTILASE_ASP; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.

```

```

DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW SIGNAL; Plasmid.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1870
FT PROPEP 1871 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT SITE 1867 1871
FT MOD_RES 1870 1870
SQ SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match
Best Local Similarity 21.9%; Pred. No. 2.5e-09;
Matches 173; Conservative 115; Mismatches 285; Indels 218; Gaps 40;

QY 17 SSSSEILQKQYIVQLHPNSHTAKTFASKFDWHLFSLQEAFLGVVEEBEESPSSLLSYG 76
Db 115 SSTAELIQ-----ETNKVIAAQ-----ASVKAAVEQVTTQTAGE---SYG 151
QY 77 SAISGFAQLTESAEAILRYSPEVAVRPHVLQVQTYYSYKFLGLDGFNS-----GVW 131
Db 152 YVANGFSTKVRVDIPKLG---QIAGVK-----TVTLAKVYPTDAKNSMANVAQVW 201
QY 132 SKSRF-GQGTIIIGVLDTGVMPESPF---DTGM---PSIPKWKKGICQEGESFSSSCN 184
Db 202 SNYKKEGTVVSVSDIGDTHDKMRLSDDKVKLTKSDVERKFTDTAKHGRYNSK--- 258
QY 185 RKLIGARFFIRGHRVANSPEESPMPREYISARDST---GHGHTASTVGGSSYSMAVNL 241
Db 259 -----VPYGFNVADND-----TITDDTVDEQGHMVAGIIGAN----- 292
QY 242 GNGAGVAR---GMAPGAHIAVYKWFNG-----CYSSDILAAIDVAIQKVDVLSLGL 293
Db 293 GTGDDPAKSVVGVAPAEQLAMKV-FTNSDTSATGSSSTLVSALEDSAKIGADVLNLSLG 351
QY 294 GFP--IPLYDDTIAIGTFRAMEGISVICAAGNNGPIESS-----VA 333
Db 352 SDSGNQTLDEPDLA-AVQNANESGTAAVISAGNSGTSGSATGKGVNCKDYGLQDNMVGTP 410
QY 334 NTAPVWSTIGAGTLDKRPFAVRLANGKLLYGEISLYPKGKIKNAGREV-----EVIYVT 387
Db 411 GTSRGATTVASAEIVTQAVTIDGT---GLQLGPGT-IQLSSNDFTSGFDQKKFVV 466
QY 388 GQDKGSEFLRGSLF--REEIRGMVCDRGVNGRSEKGEAVKBAAGVAMILANTEINQE 445
Db 467 KDAAGN--LSKGALADYTADAKGKIAIVKRGELSFDDKQKYAQAAGAGLIIVNN----- 519
QY 446 EDSIDVHLLPATLIGYT-----ESVLLKAYVN-ATVVKPARIIFGGVIGESRAP-- 494
Db 520 ---DGTATPVTSMALTTFTFGLSSVTGKLVDMVTAHPDDSL---GVKIALTLVFNQ 572
QY 495 -----EVAQSEARGPSLANPSILKPDMPAGVNIIAAWPQNLGPTGLPYDSRRVNFVMS 549
Db 573 KYTEKMSDFTSYGP-VSNLS-FKPDITAPGGINWSTQNNN-----GYTNMS 617
QY 550 GTSMSCHVSGITALLIRSAVFNWS-----PAAIKSALMTTADLYD----- 589
Db 618 GTSMASPFIAQSQALLKQALNNKNNPFYAYKQKLTALTFDLKTVEMTAQPINNDYN 677
QY 590 -----RQGAIKDGNKPAQVFAIGAGHVNPQKAINPGLVYNIQPVYITYLCT--LGF 640
Db 678 NVIVSPRQAGLVDVKA---AIDALEKNPSTVAENGYPVAVELKDFSTDKTFLTF 732
QY 641 TRSDILAITEK---NVSCNGILRNKPGFSLNYPISAVIFKRGKTTMTIRVTNVGSPNS 697
Db 733 TNSTTHELTQMSNTDTNAVYT-----SATDPNSGLYDK-----KIDGAIAKAGS--- 779
QY 698 IYSVNVKAPEG 708
Db 780 ----NITVPAG 786

```

RESULT 11
SUBT_BACSA STANDARD; PRT; 381 AA.
AC P00783;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Subtilisin amylosacchariticus precursor (EC 3.4.21.62).
GN APR.
OS Bacillus subtilis var. amylosacchariticus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89008194; PubMed=3139650;
RA Yoshimoto T., Oyama H., Honda T., Tone H., Takeshita T.,
RA Kamiyama T., Tezuru D.;
RT "Cloning and expression of subtilisin amylosacchariticus gene.";
RL J. Biochem. 103:1060-1065(1988).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=72266687; PubMed=4560201;
RA Markland F.S., Kurihara M., Smith E.L.;
RT "Subtilisin Amylosacchariticus. II. Isolation and sequence of the
tryptic and cyanogen bromide peptides.";
RL J. Biol. Chem. 247:5602-5618(1972).
RN [3]
RP SEQUENCE OF 107-381.
RX MEDLINE=72266688; PubMed=5055784;
RA Kurihara M., Markland F.S., Smith E.L.;
RT "Subtilisin Amylosacchariticus. 3. Isolation and sequence of the
chymotryptic peptides and the complete amino acid sequence.";
RL J. Biol. Chem. 247:5619-5631(1972).
CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
for peptide bonds, and a preference for a large uncharged residue
in p1. Hydrolyzes peptide amides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
IS NOT NECESSARY FOR NORMAL SPOULATION.
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; D00264; BAA00186.1; -;
DR PIR; A41448; SUESS.
DR HSP; P04189; 1SCJ.
DR MEROPS; S08.042; -;
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal.
FT SIGNAL 30
FT SIGNAL 31
FT PROPEP 106
FT CHAIN 107 381
FT ACT_SITE 138 138
FT ACT_SITE 170 170
FT ACT_SITE 327 327
FT ACT_SITE 191 191
FT CONFLICT 191

FT CONFLICT 365 365 N -> D (IN REF. 2).
SQ SEQUENCE 381 AA; 39467 MW; 2251BADE22B4824F CRC64;
Query Match 6.3%; Score 255; DB 1; Length 381;
Best Local Similarity 21.0%; Pred. No. 3.9e-10;
Matches 129; Conservative 61; Mismatches 176; Indels 248; Gaps 20;
QY 6 FFLCIIFLLFCSSSSSEILQQTIVIVQLHNSSTAKTFASKFDWHLSFLOEAVLGVVEEEE 65
DB 12 FALTLLFTWAFSNMS-----AQAAGKSTKKYIVGFQTKSAMSSAKKDKVISEK 62
QY 66 EPSRLLYSGSAIEGFAAQLTESEAEILRYSPVVAVRDPHVLOVQVTTYKFLGLDGF 125
DB 63 GGVKQKQFKY---VNAAAATLDEKAVKELKPKSPVAVVEEDHIAH-EYAQSVPY-GISQI 117
QY 126 GNSGVWSKSRFGQGTIIIGVLDTCVWPESPFSDDTGNPSIPRWKGTICQEGESPPSSSCNR 185
DB 118 KAPALHSQGYTGSNVKVAVIDSGIDSSHPDLNVRG----- 152
QY 186 KLIGARFFIRGHRVANSPEESPNMPREYISARDSTGHGTHTASTVGGSSVSMANVLNGA 245
DB 153 ---GASF-----VPSETNPYQDGSSEHGTAVGTIA-----ALNNSI 185
QY 246 GVARGMAPGAHIAVYKVMFNGC-YSSDILAAIDVAIQDVVLSLGLGFPPIPLYDDTI 304
DB 186 GVL-GVSPSASLYAVKVLDSGTSGQYSWIINGLEWAI SNMMDVINSLGG---PSGSTAL 241
QY 305 AIGTFRAMEGIGSVICAAANGNPPIESSVANTAPWSTIGAGLDRFPVAVRLANGKLLY 364
DB 242 KTVVDKAVSSGIVVAAAAGNEGSSGS-----STVG----- 272
QY 365 GESLYPCKGIGKNAGREVEIVYTGDKSEFCLRGSLPREIRGKVICDRGVNGHSEK 424
DB 273 ---YPAK----- 276
QY 425 EAVKEAGGVAMILANTEINOEDSIDVHLIPATL-IGYTESVLLKAYVNAVTKPKARIIF 483
DB 277 -----YPSIIAVGVNS----- 288
QY 484 GGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAFGVNIIAWPNLGTGLPYDSRRV 543
DB 289 -----SNQRAFSAGSEL-----DVMAFGVSI-----QSTLPGG----- 318
QY 544 NFTVMSGTSCHPVSGITALIRSAVNPNSPAKSAKMTADLYDRQGAIKDGNKPKG 603
DB 319 TYGAYNGTSMATPHVAGAAALILSKPTWTNAQVRDLSEATYL-----GNS--- 366
QY 604 VFAIGAGHVNPQKA 617
DB 367 -FYGKGLINVQAA 379

RESULT 12

SUBN_BACNA STANDARD; PRT; 381 AA.
ID SUBN_BACNA
AC P35835;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Subtilisin NAT precursor (EC 3.4.21.62).
GN APRN.
OS Bacillus subtilis var. natto.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86029;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NC2-1;
RX MEDLINE=93113095; PubMed=1369081;
RA Nakamura T., Yamagata Y., Ichishima E.;
RT "Nucleotide sequence of the subtilisin NAT gene, aprN, of Bacillus
subtilis (natto)."
RL Biosci. Biotechnol. Biochem. 56:1869-1871(1992).
CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,


```
FT HELIX 117 120
FT TURN 121 123
FT HELIX 125 130
FT TURN 131 131
FT TURN 135 136
FT TURN 138 142
FT STRAND 150 151
FT TURN 156 157
FT STRAND 173 182
FT STRAND 198 201
FT TURN 207 208
FT HELIX 213 226
FT STRAND 230 233
FT HELIX 242 252
FT TURN 253 255
FT STRAND 257 258
FT TURN 267 268
FT TURN 272 275
FT STRAND 279 279
FT STRAND 282 285
FT STRAND 291 291
FT TURN 299 300
FT STRAND 303 306
FT STRAND 310 314
FT TURN 315 317
FT STRAND 318 322
FT HELIX 325 342
FT TURN 344 345
FT HELIX 348 358
FT HELIX 375 380
SQ SEQUENCE 380 AA; 539EA7271B6682C CRC64;

Query Match 6.2%; Score 248; DB 1; Length 380;
Best Local Similarity 19.9%; Pred. No. 1.2e-09;
Matches 123; Conservative 68; Mismatches 164; Indels 262; Gaps 22;

QY 10 IIFLLFSSSEILQ--KQYIVOLHPNSETAKTFASKDFDHLSPFQEAFLGVVEEEEP 67
DB 15 LISVASSSSASAEAKEXILGFN-EQAVSEFVEQE-----ANDEVAILSEEEV 67

QY 68 SSRLSYSGSAIEGFAAQLTESAEILRYSPYAVRPHVLQVQTYGYKFLGDGFON 127
DB 68 EIELLHEF-ETIPVLSVELSPEDVDALDPAISYIEED-AEVTTMAQSVPMGSRVQA 124

QY 128 SGWKSRRFQGGIIGVLTGVPWSPSPDDTGMPSPKWKIGICOGESFSSSSCNRL 187
DB 125 PAAHNRGLTSGVKVAVLDTGI-----STHPDLNI 154

QY 188 IGARFFIRGHRVANSPEESPMMPREYISARDSTGHGTHASTVGGSSVSMANVLNGAGV 247
DB 155 RGGASFV-----PGEP-----STQDGNHGHVAGTIA-----ALANNSIGV 190

QY 248 ARGMAPGAHIAVKVCFWNCYS-SDILAAIDVAIQKVDVLSLSLGGFPPIPLYDDTIAI 306
DB 191 L-GVAPNAELYAVKVLGASGSGSVSLAQGLEWAGNGMHVANLSGS---PSPATLEQ 246

QY 307 GTFRAMEGTSVTCAGNNGPIESSVANTAPWSTTGAGTLDRFPVAVRLANGKLLYGE 366
DB 247 AVNSATSRGVLVVAASGNS-----GAGSIS--YPA--RYAN----- 278

QY 367 SLIPGKIKNAGEVEVIYTGDKSGSEFCLRGLSPREIRGNWVICDRGVNRSKGEA 426
DB 279 ----- 278

QY 427 VKEAGGVAMILANTEINOEDSIDVHLLPATLIGTYESVLLKAYVNATVKPKARIIFGCT 486
DB 279 -----AMAVGATDQNNR----- 291

QY 487 VIGESRAPEVAQFSARGPSLANPSILKPDWIAFGVNNIIAWPONLGTGLPYDSRRVNF 546
DB 292 -----ASFQYAGL-----DIVAGVNVQSTYP-----GSTVAS----- 321

QY 547 VMSGTNSCPHVSIGITALIRSAYPNWSPPAIKALMTTA-----DLYDRQKAIKDGKN 600
```

```
DB 322 -LNGTSMATPHVAGAAALVQKPNPSNSVQIRNHLKNTATSLGSTNLY----- 368
QY 601 PAGVFAIGAGHVNPOKA 617
DB 369 -----GSLVNAEAA 378

RESULT 15
ELVA_BACCS STANDARD; PRT; 380 AA.
AC P41362;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-).
OS Bacillus clausii.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=79880;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RX MEDLINE=93043753; PubMed=1368952;
RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S.,
  Aono R., Horikoshi K.;
RT "Molecular cloning, nucleotide sequence, and expression of the
  structural gene for alkaline serine protease from alkaliphilic
  Bacillus sp. 221.";
RL Biosci. Biotechnol. Biochem. 56:1455-1460(1992).
RN [2]
RP SEQUENCE OF 112-129.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RA Horikoshi K.;
RL (In) Horikoshi K. (eds.);
RL Microorganisms in alkaline environments, pp.187-194, VCH,
  Weinheim (1991).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.ebi-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC
CC EMBL; S48754; AAC60420.1; -
CC EMBL; D13157; BAA02442.1; -
CC EMBL; A26817; CAA01836.1; -
CC EMBL; A22550; CAA01611.1; -
CC HSSP; P29600; 1GCI.
CC MEROPS; S08.103; -
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PRO0723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_ASP; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 111
FT CHAIN 112 380
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 380 AA; 3826 MW; 5F73ABC6D5B6831 CRC64;

Query Match 6.2%; Score 248; DB 1; Length 380;
Best Local Similarity 19.9%; Pred. No. 1.2e-09;
Matches 123; Conservative 68; Mismatches 164; Indels 262; Gaps 22;

QY 10 IIFLLFSSSEILQ--KQYIVOLHPNSETAKTFASKDFDHLSPFQEAFLGVVEEEEP 67
```

Db 15 LISVAFSSSIAGAAEKEKYLIGN-EQEAUSEFEQVE-----ANDEVAILSEEEV 67
QY 68 SSRLLYSYGAIEGFAAQTESAEILRYSPVAVRDPDHLQVQTTYSYKFLGLDGFN 127
Db 68 EIELLHEF-ETIPVLSVELSPEDVDALEDPAISYIEED--AEVTTMAQSPVWGISRYQA 124
QY 128 SGVWSKSRFGQGTIIGVLDTGVPWSPSPDDTGMPISIRKWKGICQEGESPSSSCNPKL 187
Db 125 PAHNRGLTSGVKVAVLDTGI-----STHPDLNI 154
QY 188 IGARFFIRGHRVANSPEESPNNPREYISARDSTGHGHTASTVGGSSVSMANVLGNGAGV 247
Db 155 RRGASFV-----PGEP-----STQDNGHGHVAGTIA-----ALNNSIGV 190
QY 248 ARGMAFGAHYKVCWFNGCYS-SDILAAIDVAIQDKVDLSLSLGFPPIPLYDDTIAI 306
Db 191 L-GVAPSAELYAVKVLGASGSGSVSSIAQGLEWAGNNGMHVANLSLGS---PSPSATLEQ 246
QY 307 GTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTIDRRFPVAVRLANGKLLYE 366
Db 247 AVNSATSRGVLVVAASGNS-----GAGSTS--YEA--RYAN----- 278
QY 367 SLYPGKGIKNAGREVEVIVVTGDKGSEFCLRGSLPREIRGKMWICDRGVNGRSEKGEA 426
Db 279 ----- 278
QY 427 VKEAGGVAMILANTEINQEEEDSIDVHLLPATLIGYTESVLLKAYVYNAIVKPKARIIFGOT 486
Db 279 -----AMAVGATDQNNR----- 291
QY 487 VIGRSRAPEVAQPSARGPSLANSILKPDMIAPGUNIIAAMPQNLGPTGLPYDSRVNFT 546
Db 292 -----ASFSQYAGL-----DIVAPGVNVQSTYP-----GSTYAS----- 321
QY 547 VMSGTWSGCPHVSGITALIRSAVPNWSPAKISALMTTA-----DLYDROGKAIDGK 600
Db 322 -LNGTSMATPHVAGAAALVKQKNPWSNVQIRNHLKNTATSLGSTNLY----- 368
QY 601 PAGVFAIGAGHVNPOKA 617
Db 369 -----GSGLVNAEAA 378

Search completed: February 2, 2004, 09:10:32
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:07:47 ; Search time 41 Seconds
(without alignments)
4877.824 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEPKPFLLCIIFLLCSRSS.....NSHNLQVRSPISVTLKTN 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4018	100.0	775	064495	O64495 arabidopsis
2	1742	43.4	754	092UF6	Q92UF6 arabidopsis
3	1700	42.3	775	09LUM3	Q9LUM3 arabidopsis
4	1698	42.3	775	09C5N5	Q9C5N5 arabidopsis
5	1684	41.9	775	09LGA0	Q9LGA0 arabidopsis
6	1679	41.8	775	093205	Q93205 lycopersico
7	1666.5	41.5	757	065351	O65351 arabidopsis
8	1656	41.2	764	094H95	Q94H95 oryza sativ
9	1653.5	41.2	757	09RW07	Q9RW07 arabidopsis
10	1649.5	41.1	773	09LSE2	Q9LSE2 oryza sativ
11	1642.5	40.9	780	09FLI4	Q9FLI4 arabidopsis
12	1640.5	40.8	746	093907	Q93907 arabidopsis
13	1611	40.1	766	093204	P93204 lycopersico
14	1600	39.8	777	09LVJ1	Q9LVJ1 arabidopsis
15	1589	39.5	764	094607	Q94607 arabidopsis
16	1572.5	39.1	754	065834	O65834 lycopersico

17	1562	38.9	756	10	Q9LPD1	Q9LPD1 arabidopsis
18	1542	38.4	754	10	O65835	O65835 lycopersico
19	1534.5	38.2	745	10	O04678	O04678 lycopersico
20	1521	37.9	747	10	O82007	O82007 lycopersico
21	1513.5	37.7	747	10	O65836	O65836 lycopersico
22	1509.5	37.6	747	10	O65836	O65836 lycopersico
23	1504	37.4	746	10	Q9LWA3	Q9LWA3 lycopersico
24	1503.5	37.4	743	10	Q9LWA4	Q9LWA4 lycopersico
25	1490	37.1	745	10	O96478	O96478 lycopersico
26	1492.5	36.9	754	10	Q8H047	Q8H047 oryza sativ
27	1473	36.7	772	10	Q9ZSP5	Q9ZSP5 arabidopsis
28	1463.5	36.4	759	10	O8RVC2	O8RVC2 oryza sativ
29	1445.5	36.0	758	10	O82440	O82440 arabidopsis
30	1440.5	35.9	755	10	O8S896	O8S896 arabidopsis
31	1413	35.2	666	10	Q9ZK46	Q9ZK46 lycopersico
32	1385.5	34.0	778	10	O9AX30	O9AX30 oryza sativ
33	1365	34.0	760	10	Q9FJF3	Q9FJF3 arabidopsis
34	1348.5	33.6	581	10	O8W554	O8W554 arabidopsis
35	1328.5	33.1	756	10	O8GXU1	O8GXU1 arabidopsis
36	1311.5	32.6	789	10	O9LRF2	O9LRF2 oryza sativ
37	1310.5	32.6	762	10	O8H4X8	O8H4X8 oryza sativ
38	1309	32.6	791	10	O9FK76	O9FK76 arabidopsis
39	1282	31.9	761	10	Q38708	Q38708 alnus glut
40	1279	31.8	753	10	O8S1N4	O8S1N4 oryza sativ
41	1272.5	31.7	757	10	O8L4T6	O8L4T6 oryza sativ
42	1272	31.7	774	10	O9MAP7	O9MAP7 arabidopsis
43	1271.5	31.6	756	10	O9ZSB0	O9ZSB0 arabidopsis
44	1248.5	31.1	778	10	O8LIH3	O8LIH3 oryza sativ
45	1244.5	31.0	760	10	O8S1N3	O8S1N3 oryza sativ

ALIGNMENTS

RESULT 1
O64495 PRELIMINARY; PRT; 775 AA.
AC O64495
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE F20D22.12 protein.
GN F20D22.12.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Euphorbiales; Brassicales; Brassicaceae; Arabidopsi.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Vyotskaia V.S., Osborne B.I., Schwartz J.R., Toriumi M., Kwan A.,
Yu G., Oji, O., Liu S., Li J., Hoang L., Araujo R., Au M., Brendel V.,
Buehler E., Conway A.B., Conway A.R., Dewar K., Feng J., Kim C.,
Kurtz D., Li Y., Palm C.J., Shinn P., Sun H., Davis R.W., Ecker J.R.,
Federapfel N.A., Theologis A.;
RA "Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence";
RT Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Yu G., Oji, O., Liu S., Li J., Hoang L., Araujo R., Au M., Brendel V.,
Buehler E., Conway A.B., Conway A.R., Dewar K., Feng J., Kim C.,
Kurtz D., Li Y., Palm C.J., Shinn P., Sun H., Davis R.W., Ecker J.R.,
Federapfel N.A., Theologis A.;
RA "Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence";
RT Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC002411; AAC16749.1;
DR HSP; P00782; 2SBT.
DR MEROPS; S08.084;
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILIN.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.

SQ SEQUENCE 775 AA; 83775 MW; 7740B20397C7C211 CRC64;

Query Match 100.0%; Score 4018; DB 10; Length 775;
 Best Local Similarity 100.0%; Pred. No. 2.5e-265;
 Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKPPFFLCIIIFLLFCSSSEILQKQTVIVOLHNSTAKTFASKDWHLSFLQEAFLGV 60
 DB 1 MEKPPFFLCIIIFLLFCSSSEILQKQTVIVOLHNSTAKTFASKDWHLSFLQEAFLGV 60

QY 61 EEEBEPSSRLLYSGSAIEGFAAQLTESEAEILRYSPVAVRPHVQLQVQTVSYKPL 120
 DB 61 EEEBEPSSRLLYSGSAIEGFAAQLTESEAEILRYSPVAVRPHVQLQVQTVSYKPL 120

QY 121 GLDGFNSGVWKSFGGTTIGVLDGVWPESSEDDTGMPISPRKWKGIQCEGESFSS 180
 DB 121 GLDGFNSGVWKSFGGTTIGVLDGVWPESSEDDTGMPISPRKWKGIQCEGESFSS 180

QY 181 SSCNRKLIGARFFIRHGVANSPESPNMPREYISARDSTGCHTHTASTVGGSSVMANV 240
 DB 181 SSCNRKLIGARFFIRHGVANSPESPNMPREYISARDSTGCHTHTASTVGGSSVMANV 240

QY 241 LKNGAGVARGMAPGHAIVKVCNFCYSDIILAAIDVAIQKVDVLSLGGPPIPL 300
 DB 241 LKNGAGVARGMAPGHAIVKVCNFCYSDIILAAIDVAIQKVDVLSLGGPPIPL 300

QY 301 DDTIAIGTFRAMEGSIIVICAGNNGPIESSVANTAPWGTIGAGTLDRRFAVVLNGLANG 360
 DB 301 DDTIAIGTFRAMEGSIIVICAGNNGPIESSVANTAPWGTIGAGTLDRRFAVVLNGLANG 360

QY 361 KLYGESLYPGKIKNAGREVEVIVYVGDKSGSEFFCLRGSLPREEIRGKMWICDRGVNR 420
 DB 361 KLYGESLYPGKIKNAGREVEVIVYVGDKSGSEFFCLRGSLPREEIRGKMWICDRGVNR 420

QY 421 SEKGAIVKAGGAVAMILANTINOEBSIDVHLLPATLIGYTESVLLKAYVNAVTKPKAR 480
 DB 421 SEKGAIVKAGGAVAMILANTINOEBSIDVHLLPATLIGYTESVLLKAYVNAVTKPKAR 480

QY 481 IIFGTVIGRSAPVAQFARGPSLANPSILKPDMIAPGVNTIAWPNQNLGTPGLPYDS 540
 DB 481 IIFGTVIGRSAPVAQFARGPSLANPSILKPDMIAPGVNTIAWPNQNLGTPGLPYDS 540

QY 541 RVNFTVMSGTSMSCPHVSIGITAIRSAYPNWSAAIKSALMTTADLYDROGKAIKDGK 600
 DB 541 RVNFTVMSGTSMSCPHVSIGITAIRSAYPNWSAAIKSALMTTADLYDROGKAIKDGK 600

QY 601 PAGVFAIGAGHVPKAINPGLVNIOPVDYITVCTLGTFRSDILAITHKNVSCNGILR 660
 DB 601 PAGVFAIGAGHVPKAINPGLVNIOPVDYITVCTLGTFRSDILAITHKNVSCNGILR 660

QY 661 KNPGLSLNYPSTAVIFKRGKTEMITRRTVNVGSPNSIYVNVKAPGKIVNPKRLV 720
 DB 661 KNPGLSLNYPSTAVIFKRGKTEMITRRTVNVGSPNSIYVNVKAPGKIVNPKRLV 720

QY 721 KHVDTQLSYRVNVLKXNRGKVASPAQGLTWNHNLQVRSPISVTLKTN 775
 DB 721 KHVDTQLSYRVNVLKXNRGKVASPAQGLTWNHNLQVRSPISVTLKTN 775

RESULT 2
 Q92UF6
 ID Q92UF6 PRELIMINARY; PRT; 754 AA.
 AC Q92UF6;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
 DE Putative serine protease (Putative subtilisin serine protease).
 GN A72G05920.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;

[1] SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White C., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:761-768(1999).
 [2] SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3] SEQUENCE FROM N.A.
 RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
 RA Bowers L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
 RA Jones T., Kamiya A., Karir-Neumann G., Kawai J., Kim C., Koesema E.,
 RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
 RA Palm C.J., Sakurai T., Satou M., Seki M., Shim P., Southwick A.,
 RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full Length cDNA of Gene T6P5.12/At2G05920 (GI:4006827).";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 [4] SEQUENCE FROM N.A.
 RP Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,
 RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C.,
 RA Wu H.C., Yu G., Yuan S., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W.,
 RA Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; ACC05970; AAC95169.1; -
 DR EMBL; AY035090; AAK9595.1; -
 DR EMBL; AY142613; AANJ3182.1; -
 DR HSP; Q99405; 1MPT.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 2.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00840; PA; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Protease.
 SQ SEQUENCE 754 AA; 80015 MW; F8CCG3F11578CE7 CRC64;

Query Match 43.4%; Score 1742; DB 10; Length 754;
 Best Local Similarity 48.4%; Pred. No. 3.9e-110;
 Matches 370; Conservative 118; Mismatches 237; Indels 40; Gaps 16;

QY 11 IFLLFCSSSEILQKQTVIVOLHNSTAKTFASKDWHLSFLQEAFLGVVEEPESSR 70
 DB 18 LFLLETTA----KKTIIIRVN-HSDKPEFUTHHEDWYTSQLS-----ESS 59

QY 71 LLYVGSAGEGFAAQLTESEAE-ILRYSPEVAVRPHVQLQVQTVSYKFLGLDG-FGNS 128
 DB 60 LLYVTTTFHGFSAVLDSTADSLSSNSILDFEDPLTYLHTTTPPEPLGNSFGVH 119

QY 129 GWMSKSRFGQTTIGVLDGTGVWPESFDDTGMPISPRKWKGIQCEGESFSSSCKRUI 188
 DB 120 DLGSSS---NGVIIGVLDGTGVWPESRFDGTDNPEIPSKNKGCESSGSDPDLCKKUI 176

QY 189 GARFTIRHVRVNSPEESPNNPREYISARDSTGCHTHTASTVGGSSVMANVLCNGAGVA 248
 DB 189 GARFTIRHVRVNSPEESPNNPREYISARDSTGCHTHTASTVGGSSVMANVLCNGAGVA 248

Db 177 GARSFSGFOMASGGGFSK--RESVSPRDVGHGHTSTTAAGSAVRNASFLGVAAGTA 234
 QY 249 RGMAPGAHIAVYKVCWFGNCGYSSDILAAIDVAIQKVDVLSLGGFFPIPLYDDTIAICT 308
 Db 235 RGMATRARVATYKVCWSTCFGSDILAAANDRAILLDGVVLSLGGGSAFYRDIIAIGA 294
 QY 309 FRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDLRFPFPAVRRIANGLLYGSSL 368
 Db 295 FSAMERGVSFVSCAGNSGPTASVANVAPWMTVAGTLDLRFPFPAVRRIANGLLYGSSL 354
 QY 369 YPGKGIKNAGREVEVLYVGGKGEFCLRGSI,PREEIRGMKVICDRGVNGSESEGAVK 428
 Db 355 YSGVGM--GTXPLELYNKGSSSNLCUPGLSDSSIVRGKIWCVRGYNARVEKGVYR 412
 QY 429 EAGGVAMILANTEINOEDSIDVHLLPATLIGVTSVLKAYVATVVKPARIIEGGTVI 488
 Db 413 DAGGLWIAWNTAASGEELVANSHLLPAIVAGKKTGDLLEVVYKSDSKETALLVFKGIVL 472
 QY 489 GRSRAPEVAQFARGPSLANPSILKPDMIAPGNIIAAMPONLGTGLPYDSRRYNTVM 548
 Db 473 DVKPSVWAFAFSRGNTVPEILKPDVIGPGVNIILAGWSDAIGPTGLDKDSRRTOFNIM 532
 QY 549 SGTSMSCPHVSGITALIRASYPNWSPAAIKSALMTADLYDQKAIKDG--NKPAGVFA 606
 Db 533 SGTSMSCPHISGLAGLLKAAHPWSPSAIKSALMTAYVLDNTNAPLHDAADNSLSNRYA 592
 QY 607 IGAGHYNPOKAIPNGLVNIQPDVITYLTCTLGFTSRDILAIHK--NVSCNGILKKNPGF 665
 Db 593 HSGSHVDPOKALSPGLVYDITSTEEYIRFLCSLDYTVDHIVAIVKPSVNCCKF--SDPG- 650
 QY 666 SLNYPGSIIVFKRGKTKTEMITRRVTNVGSPNSIYVNVKAPGKIVIVNPKRLVFKHVDQ 725
 Db 651 QLNYPGFSVLF--GKRVVRYTRETVNVAASSVYKVTYNGAFSGISVKPSKLSFKSVGE 709
 QY 726 TISYRVFVFLKKNRGKGVASFAQ--GQLTVNWSHNLMOVRSPIS 769
 Db 710 KKRYTVFVSKK---GVSMNKAEFGSITWSNPOH---EVRSPVA 748

RESULT 3

Q9LUM3 ID Q9LUM3 PRELIMINARY; PRT; 775 AA.
 AC Q9LUM3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Subtilisin proteinase-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 EX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 RT clones";
 RL DNA Res. 7:131-135(2000).
 DR ENBL; AB022220; BAB01030.1; --
 DR HSP; Q99405; IMPT.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 2.
 DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS50840; PA; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 SQ SEQUENCE 775 AA; 82584 MW; CC909DE4DABDE0E8 CRC64;
 Query Match 42.3%; Score 1700; DB 10; Length 775;
 Best Local Similarity 45.0%; Pred. No. 3e-107;
 Matches 356; Conservative 134; Mismatches 257; Indels 44; Gaps 15;
 QY 6 FFLCIIFLLSCSSSEILQKQ--TVIVQLHPNSETAKTFASKFDWHLSPQLQEAVLGVEE 62
 Db 3 FFYFFFLLLTSFSSSASSSSNSLTIVVHDHAKPS-IPPTHFWYTSLSLIT----- 56
 QY 63 EEEPPSRLLYSGSAIEGFAAQLTSEAEILRYSPEWAVRPHVLQVQTYTSYKFLGL 122
 Db 57 --SSPPS-IITHYDTVPHGFSARLTQSQASQLLDHPHVISVIPEQVRHLHTTSPFLGL 113
 QY 123 DGFNSGVWSKSRFGQGTIIIGVLDTGWPSPSPDDTGMPISIRKWKGIQOEGESSSSS 182
 Db 114 RSTDKAGLLBEESDFGSLVIGVDTGVWPERPFDRLGPGVPIKWKGCQCIASQDFPESA 173
 QY 183 CNRKLIGARFFIRGHRVANSPEESPNNPREYIARDSTGHGHTASTVGGSSYSVANVLG 242
 Db 174 CNRKLIGARFFCGGYEATNG---XMETTEFRPRSDGHTHTASISAGRYVFPASTLG 230
 QY 243 NGAGVARGMAPGAHIAVYKVCWFGNCGYSSDILAAIDVAIQKVDVLSLGGFFPIPLYDD 302
 Db 231 YAHGVAAGMAPKARLAAYKVCWNSGCYSDILAAFDATAVDGVDVLSLGGVVPVYLD 290
 QY 303 TIAIGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDLRFPFPAVRRIANGL 362
 Db 291 AIAIGAFGAIDRGIFFVSASAGNGGPGALITVTVAPWMTTVAGTIDRDPANVKGNGKM 350
 QY 363 LYGESLYPGKGIKNAGREVEVY---VTGSD--KGSFCLRGSLPREIRCKMVICDRGVN 418
 Db 351 ISGVSVYGGPGL--DPCRMVPLVYGGSLGGDGYSSLCLEGSLDPNLVKGI VCLDRGIN 409
 QY 419 GRSEKGEAVKEAGVAMILANTEINOEDSIDVHLLPATLIGVTSVLKAYVATVVK-- 476
 Db 410 SRATKGEIVKNGGLGMIANGVPDGEGLVADCHVLPATSVGASGGDEIRRYISESKSR 469
 QY 477 ----PKARIIFGGTVIGRSRAPEVAQFARGPSLANPSILKPDMIAPGNIIAAMPONLG 532
 Db 470 SSKEPTATIVFKGTRLGIRPAPVVASFSARGPNPETPEILKPDVIAPIAGNLIAAWPDRIG 529
 QY 533 PTGLPYDSRRYNTVMSTGSMSCPHVSGITALIRASYPNWSPAAIKSALMTADLYDROG 592
 Db 530 PSQVTSNRETEFNILSGTSMACPHVSGLAALIKAAHPDMSPAAIRSALITATTAYVDNSG 589
 QY 593 KAIKD--GNKPAGVPAIGAGHYNPOKAIPNGLVNIQPDVITYLTCTLGFTSRDILAIT 649
 Db 590 EPMMDSTGN--TSSVMDYGGSHVHPTRKAMPDGLVYDITSYDYINFLCNSTYTTNTVIT 648
 QY 650 HKVSCNGILRKNPGFSLNYPSTAVIFKRGKTKTEMIT---RRVTNVGSPNSIYSVANVKAP 706
 Db 649 RRQACDCGARRAGHVGNLVNYPFSVVFQYQGESKMSHTFIRVTNVGSDSVYEIKIRPP 708
 QY 707 EGKIVIVNPKRLVFKHVDQTLSYRVNMFVLKKNRGKGVASFA----QGQLTVNWSHNLMO 762
 Db 709 RGTIVTVEPEKLSFRRVGQKLS---FVVRVKTEVKLSPGATNVTETGHIWSDG---KR 761
 QY 763 RVRSPISVTLK 773
 Db 762 NVTSPLVVTIQ 772

RESULT 4

Q9CSN5 ID Q9CSN5 PRELIMINARY; PRT; 775 AA.
 AC Q9CSN5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical 82.5 kDa protein.

```

GN MLN21.2/AT3G14240.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldemich A.D., Lee J.M., Quach H.B., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene MLN21.2/AT3G14240 (GI:9279572).";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF360129; AAK25839.1; -.
DR HSSP; O99405; 1MPT.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Hypothetical protein.
KW VARIANT
FT VARIANT
SQ SEQUENCE 775 AA; 82524 MW; 52FA2C92B1C652C8 CRC64;

Query Match 42.3%; Score 1698; DB 10; Length 775;
Best Local Similarity 45.0%; Pred. No. 4.1e-107;
Matches 356; Conservative 134; Mismatches 257; Indels 44; Gaps 15;

QY 6 FFLCIIIFLLPCSSSSSEILQKQ---TYIVQLHPNSETAKTFASKFDWHLSTFLOEAVLGVEE 62
DB 3 FFFYFFELLTSSPSSASSNSLTIVVVDHEAKPS-IPFTHFHWYTSLSASL-----56

QY 63 EEEPSRLLYSGSAIEGFAAQLTSEAEILRYSPPEVAVRDPDHLVQVQVTSYKFLGL 122
DB 57 --SSPPS-IHTYNTVHGFSAARLTSDASQLLDHPHVISIPEQVRLHHTRSPEFLGL 113

QY 123 DGFNGSWGSKRFGQGTIIIGVLDGTWGPSPFDDTGMPSIPRKWKIGIQEGESPPSSS 182
DB 114 RSTDKAGLLEESDFGSLVIGVDTGWMPERPSPDFDRGLGVPFIKWKGCQIASQDFPESA 173

QY 183 CNRKLIGARFIRGHRVANSPEESPNNPREYISARDSTGHGTHTASTVGGSSVMAVILG 242
DB 174 CNRKLIGARFIRGHRVANSPEESPNNPREYISARDSTGHGTHTASTVGGSSVMAVILG 230

QY 243 NGAGVARGMAPGAHIAVYKWFNGCYSSDILAADVAIQDKVDVLSLSLGGFFPIPLYDD 302
DB 231 YAHGVAAGMAPKARLAAYKVCWNSGCYSDILAADVAIDTAVADGVVLSVGGVVPYILD 290

QY 303 TIAIGTFRAMEGISVICCAAGNGPIESSVANTAPWVSTIGAGTLDLRRPAPVVRANGKL 362
DB 291 AIAIGARGAIDRGIFVSASAGNGPGALITVTVAPWMTTVGATIDRDPFANVILNGRM 350

QY 363 LYGESLYPGKIGKNACREVEVIY---VTGCD-KGSEFFCLRGSLPREIEISKVVICDRGVN 418
DB 351 ISGVSVYGGPGL-DPGRMYPVTVYGGSLGGDGYSSSLCEGSLDPLNVKGVLCVLCRGIN 409

QY 419 GRGKGAIVEAGVAGVAMILANTINOEDSIDVELLPATLIGVTSVILKAVYNAVTK--476
DB 410 SRATKGEIVKNGGLGMLIANGVDFGEGLVADCHVLPATSVGASGGDEIRVYSSSKR 469

QY 477 ----PKARIIPGTVIGRSRAPVAFSARGPSLANPSILKPDMPAPIAGVNIITAAVNLG 532
DB 470 SSXKPTATIVFKGTRLGIRAPVAVSARGPNERPEILKPDVIAAPGLNLAAMPDRIG 529

QY 533 PTGLPYDSRRVNTVMSGTSCHVSGITALLTRSAYPNWSPAAIKSAIMTTADLYDROG 592
DB 530 PSGYTSDNRRTFENILSGTSMACPHVSGLAALLKAAHPDMSPAAIRSALITTAITYVDNSG 589

```

```

QY 593 KAIKD---GNKPAGVFAIGAGHVNPQKALNPLGLVYNIQPDVYITLCTLGFTRSDILAIT 649
DB 590 EPMWDESTGN-TSSVWDYSGVHPFKAMDPLGVLDITSVDYINFLCNSNTRTNIVTIT 648
QY 650 HKNVSCNGILRNKPGSLNPSYIAVIFKRGKTEMIT---RRVTNVSNSYSYVNVKAP 706
DB 649 RRQADCDGARRAGHVLGNLPFSVVFQQYSGSKSTHFIRTVTNVGDSDSYVEIKIRPP 708

QY 707 EGIKTVNPKRLVFKHVDCTLSYRVWFVLKKNRGKVASFA----QGQLTWVNSHNLMO 762
DB 709 RGTIVTVEPKUSFRRVGQKLS----FVVRKTEVKLSFGATNVETGHVWSDG---KR 761
QY 763 RVASPISTVLK 773
DB 762 NVTSPVLVTLQ 772

RESULT 5
QBLGAG PRELIMINARY; PRT; 775 AA.
AC QBLGAG;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE Subtilisin-like serine protease.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
annotation.";
RL Genome Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY084387; AAM0964.1; -.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease.
SQ SEQUENCE 775 AA; 82456 MW; 28E8753844348303 CRC64;

```

```

Query Match 41.9%; Score 1684; DB 10; Length 775;
Best Local Similarity 44.8%; Pred. No. 3.7e-106;
Matches 354; Conservative 133; Mismatches 260; Indels 44; Gaps 15;

QY 6 FFLCIIIFLLPCSSSSSEILQKQ---TYIVQLHPNSETAKTFASKFDWHLSTFLOEAVLGVEE 62
DB 3 FFFYFFELLTSSPSSASSNSLTIVVVDHEAKPS-IPFTHLHWTSSLASLT-----56

QY 63 EEEPSRLLYSGSAIEGFAAQLTSEAEILRYSPPEVAVRDPDHLVQVQVTSYKFLGL 122
DB 57 --SSPPS-IHTYNTVHGFSAARLTSDASQLLDHPHVISIPEQVRLHHTRSPEFLGL 113

QY 123 DGFNGSWGSKRFGQGTIIIGVLDGTWGPSPFDDTGMPSIPRKWKIGIQEGESPPSSS 182
DB 114 RSTDKAGLLEESDFGSLVIGVDTGWMPERPSPDFDRGLGVPFIKWKGCQIASQDFPESA 173

QY 183 CNRKLIGARFIRGHRVANSPEESPNNPREYISARDSTGHGTHTASTVGGSSVMAVILG 242
DB 174 CNRKLIGARFIRGHRVANSPEESPNNPREYISARDSTGHGTHTASTVGGSSVMAVILG 230

```

```

174 CNRKLVGARFFCGGYEATNG---KWNETTEFRSPRSDSGHGHTWTSASIGRYYVFPASTJG 230
243 NGAGVARGMAPGAHIAVYKVCWFNGCXSSDILAAIDVAIQDVDLVLSLSLGGPFIPLYDD 302
231 YAHGVAAGWAPKARLAAVKVCWNSCYDSDILAAFDTAVADGVDVLSLSVGGVVPVYLD 290
303 TTAICTFRAMERGISVICAAGNNGPIESSVANTAPWSTTIGAGTLDPRPPAVVRLANGKL 362
291 AIAIGAFAIDRGIFVSASAGNGGFGALTVTVNAPWMTTVGACTIDRDPANVYKLGNGXM 350
363 LYGESLPGKGIKNAGREVEVIY---VTGGD-KGSEFCFLRGSLPREBIEGRKMWICDRGVN 418
351 ISGVSVYGGPGL-DGRMYPVLYCGSLGGYGSSSJCLEGSLDPNLVTGKIVLCDRGIN 409
419 GRSEKGEAVKEAGGVAMILANTEINQEBDSIDVHLLPATLIGYTESVLLKAYVNAFVK-- 476
410 SRATKGEIVRKNGGLGMIANGVFDEGLVADCHVLPAFVSAGSAGDERRVYSESSKGR 469
477 ----PKARIIIFGGTVIGRSRAPEVAQSARGPSLANPSILKPDWIAFGVNIILAAWPQNLG 532
470 SKKHPTATTVFNGTSLGRTPAVVASFARGNPETPEILKPDVIAPLGNILAAWPDRIJ 529
533 PGLPYDERRVRNFTVMSGTSMSCPHVSGITALITRSAYPNWNSPAAIKSALMTADLYDROG 592
530 PSGVTSNDRTEFNILSGTSMACPHVSGLAALLKAAHPDWSPAIIRSALMTAYTVDNSG 589
593 KAIKD---CNKPAGVFATGAGHNPQKAINPGLVYNTQPVDYITYYJCTGIFTRSDILAIT 649
590 EPMWDESTGN-TSSVTDYGSGEVHPTRAMDPLGYDITSDYDINFLCNSNYTGTNIVTIT 648
650 HKNVSGNGLRKNPGFSLNYPISIAVIFKRGKTTWMIT---RRVTNVSGPSNIYSVNVKAP 706
649 RQACDCGARRAGHVGNLUNYPSFVVPQYGESKXSTHFTIRTVTNVGDSDSVTEIKIRPP 708
707 EGIKVINPRLVFKHVDQTLISYRVWFVYLLKKNRGGKVASFA----QGOJTWVNSHNLQJ 762
709 RGTVTVBEKLSPRVQCKLS---FVVRVKTTEVKLSPGATNVETGHWMSDG---KR 761
763 RVRSISVTLK 773
762 NVTSLVWTLQ 772

```

RESULT 6	PRELIMINARY;	PRT;	775 AA.
F93205			
ID	P93205		
AC	P93205;		
DT	01-MAY-1997 (T-EMBLrel. 03, Created)		
DT	01-MAY-1997 (T-EMBLrel. 03, Last sequence update)		
DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)		
DE	Serine protease, SBT2.		
DE	SBT2.		
GN	GN		
OS	Lycopersicon esculentum (Tomato).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Asteridae; lamids; Solanales; Solanaceae; Solanum.		
OX	NCBI_TaxId=4081;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Schaller A.;		
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=cv. VFWS;		
RC	Meichtry J.; Anrhein N.; Schaller A.;		
RT	"The gene family of subtilisin-like proteases in tomato.";		
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; X98930; CAA67430.1; -		
DR	EMBL; AJ006379; CAA07000.1; -		
DR	HSSP; Q45670; 1DBI		
DR	InterPro; IPR000183; Decarboxylase2.		
DR	InterPro; IPR003137; PA.		
DR	InterPro; IPR000209; Peptidase_S8.		


```

QY 180 SSSCNKLGARFIRGHRVANSPP-EESNMPREYISARDSTGCHGTHHTASTVGGSSVMA 238
DB 173 ASLCNKLIGARFARGYSTWGPIDES-----KERSPRDDCHGTHHSSTAAGSVVGA 228
QY 239 NVLNGAGVARGMAPGAHIAVVKVWFCNGCYSDILAAIDVAIQDKVDVLSLSLGGFFIP 298
DB 229 SLIGYASGTARGWAPRARVAVVKVWLGCGFSSDILAAIDKAIADNVNVLMSLGGGMSD 288
QY 299 LYDDTTAIGTFRAMERGISVICAAGNNGPIESVANTAPWSTIAGTILDRPFAVRIA 358
DB 289 YRDGVAIGAFAAMERGIIVSCSAGNAGPSSSLSNVAFWITTVAGTILDRPFAALATIG 348
QY 359 NGKLLGESILPGKGKKNAGREVEVIYVVGDKGS--EFLAGSLPREBEIRGKVICDRG 416
DB 349 NGKNFTGVSLFKGEALPD--KLPLFIYAGNANNGNLCMTGTILPEKVGKIVMCDRG 406
QY 417 VNGRSEKGEAVKEAGGVAMILANTEINOEDSIDVHLLPATILIGYTESVLLKAYVNAIVK 476
DB 407 INARVQGDVVKAAGGVGMILANTAANGEELVADAHLLPATTVGEKAGDIIRHYVTTDPN 466
QY 477 PKARIIFGGTVIORSRAPEVAQPSLANPSILKPDMIAPGVNIIAAMPONLGGPTGL 536
DB 467 PTASISILGTVVGKPSVVAAPFSSRGPNISITPNILKPDLIAPGVNIIAAMTGAAGPTGL 526
QY 537 PYDSRRVNFVMTGWSNCPHUSGITALIRSAVPNWSPAIKSALMTTADLYDRGKAIK 596
DB 527 ASDRRVEFNIISGTNSCPHVSGLAALKSHVPECSPAIRSAALMTTAYKYDKGKPLL 586
QY 597 D--GNKPAGVFAIGAGVMPQKAINPGLVNIQPDVYIYVLTGTFTRSDILAIHTKNVS 654
DB 587 DIATGPESTPFDHAGHVSPTTATNGLIYDLTTEDYLGFLCALNYTSQIBSRVSRNT 646
QY 655 CNGILRKNPGFS---LNPYSIAVIFKRGKTEMITRVTNVGSPNSIYSVNVKA--PEGIK 710
DB 647 CD----PSKSYSVADLNYPESFAVND--GVGAVKYTRVTSVGGAGT--YSVKVTSSETGVK 700
QY 711 VIYNPKRLFKHVDQTLISRVWVVLKKRGGKGVASFAQGLTWNHNLQVRSPISV 770
DB 701 ISVEPAVLNPKANEKKSYYVTFTVDSXPSG---SNSFGSIEWSDGKHV---VGSVAI 754
QY 771 T 771
DB 755 S 755

RESULT 10
Q8LSS2
ID Q8LSS2 PRELIMINARY; PRT; 773 AA.
AC Q8LSS2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative cucumis-in-like serine protease.
GN OSUNBA001109.20.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Kim M.M.,
RA Overton II L.L., Bera J.J., Tsitrin T., Krol M.I., Jarrahi B.B.,
RA Jin S.S., Xoo H., Zismann V., Heiao J., Blunt S., Vanaken S.S.,
RA Uterback T.F., Feldblyum T.V., Yang Q.Q., Haas B.J., Suh B.B.,
RA Peterson J.J., Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RA "Oryza sativa chromosome 10 BAC OSUNBA001109 genomic sequence.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC092388; AAN22744.1; -
DR Gramene; Q8LSS2; -.
DR InterPro; IPR003137; PA.

```

```

DR InterPro; IPR000209; Peptidase_s8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease.
SQ SEQUENCE 773 AA; 79132 MW; 71602DBE5D435DC3 CRC64;

Query Match 41.1%; Score 1649.5; DB 10; Length 773;
Best Local Similarity 47.0%; Pred. No. 8.2e-104;
Matches 361; Conservative 99; Mismatches 257; Indels 51; Gaps 15;

QY 27 TYIVOLHPNSETAKTFASKFDMHLSFLQEAVLGVVEEBEPSSRLLYSYCSALEGFAAQL 86
DB 35 TYIVFMDP-----ARLPAA---GHAHLQSLAI-----DPDRHLLYSYSAAHGFAAL 80
QY 87 TSEAEATIRYSPEVAVRDPHVLQVQVTSYKFLGLDGFNGSVMSKSRFG-----QGTI 141
DB 81 LPHILPLLRASPGVLQVWPDEVFLHTTRTFEPLGL----LSPAYQPAIHGFEATHDVV 136
QY 142 IGVLDTCVWPESFDDTGMPSTIPRKWKIGICQGESFSFSSSCNRKLIIGARFIRGHRVAN 201
DB 137 IGVLDTCVWPESFAGGDLPPPPARMKGVCEAGVDFSPVCGKILVGARSFSRGLRAAN 196
QY 202 ----SPEESPMPREXISARDSTGHGTHHTASTVGGSSVSMANVLNGAGVARGMAPGAHI 257
DB 197 GGGGGGARGVGRRGFSVARDRDRGHTHTATTAGAVANASLLGYATGTARGMAFGARV 256
QY 258 AVKVCWFNFCYSDILAAIDVAIQDKVDVLSLSLGGFFIPLYDDTTAIGTFRAMERGIS 317
DB 257 AAYKVCWPEGLGSDILAGIDAAVADGVGVLSLSLGGGAPYFRDITVAVGAFGAAAAGVF 316
QY 318 VICAGNNGPIESVANTAPWSTIAGTILDRPFAVRLANGKLLYGESILPGKIKNA 377
DB 317 VACSAGNSGSGATVANSAPWATVAGTILDRPFAVTLPTGARLAGVSLYAGSPSPR 376
QY 378 GREVEVIYVVGDKGSFELRGSIPREIRGKVIQDRGVNGRSEKGEAVKEAGGVAMIL 437
DB 377 PAMLPVLYGGGDNASRLCLPGTLDPAVRGKIVLCDRGVNARVEKAVVKAAGAGMVL 436
QY 438 ANTEINOEDSIDVHLLPATILIGYTESVLLKAYVNAIVT-----KPKARIIFGGTVIGRSR 492
DB 437 ANTAASGEELVADSHLLPAVAVGKLAGDKIREVASRRAGAGAPVAILSGGTVLGVRP 496
QY 493 APEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPONLQPTGLPYDSRVNFTVMSGTS 552
DB 497 SPVVAAPFSSRGPNVTPVPEILKPDMIAGVGVNLAGVSGVAGTGLVKDGRRTHEINISGTS 556
QY 553 MSCPHVSGITALIRSAVPNWSPAIKSALMTTADLYDRGKAIKD--GNKPAGVFAIGAG 610
DB 557 MSCPHISGVAALLKAAHFEWSPAAIKSALMTTATYVDNTNSSLRDAAGGLATPFAFGAG 616
QY 611 HVPQKAINPGLVNIQPDVYIYVLTGTFTRSDILAIHT-KNVSCNGILRKNPSPSLNY 669
DB 617 HVDQKALSPCLLYDITKDYVSFLCSLNTTTHIQTITKMSNITCPKFR--PG-DLNY 673
QY 670 PSIAVIF-KRGKTEMITRVTNVGSPNSIYSVNVKAPEGIKVIVNPKRLVFKHVDQTL 728
DB 674 PSFSVFKKSKHVMRFRREVTVNGVPMSVYVNVKVGSPASVSVKVTPAKLVFNKVGQKOR 733
QY 729 YRVWFLVKKXNRGGKVASFAQ---GQLTWNHNLQVRSPISVTLK 773
DB 734 YVVFASVTVD-----ASNAPDGFWMSSQHV---VRSPIATYWK 772

RESULT 11
Q9FLI4
ID Q9FLI4 PRELIMINARY; PRT; 780 AA.
AC Q9FLI4;
DT 01-MAR-2001 (TrEMBLrel. 15, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```

```

DE Serine protease-like protein (Putative subtilisin serine
DE protease).
GN AF551750.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
CX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT Sequence features of the regions of 1,456,315 bp covered by nineteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:41-54(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kaniya M., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Arabidopsis Full length cDNA Clones.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB010074; BAB1244.1; -.
DR EMBL; AY080831; AAL87307.1; -.
DR HSP; Q45670; IDBI.
DR MEROPS; S08.092; -.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00879; ODR_DC_2_2; 1.
DR PROSITE; PSS0840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease.
SQ SEQUENCE 780 AA; 84950 MW; BE22D41550C8BCFC CRC64;

Query Match 40.9%; Score 1642.5; DB 10; Length 780;
Best Local Similarity 43.6%; Pred. No. 2.5e-103;
Matches 342; Conservative 128; Mismatches 281; Indels 33; Gaps 11;

QY 4 KPFFLCI---IFLLFCSSSEILQKQTVIVQLHPNSETAKTFASKFDWHLFLQEAVLG 59
DB 9 KPFLFILNLFLQAEITVQISTKTKYVIHM-DKSMPLPYTNHLQWYSSKINSVTOH 67
QY 60 VEEEEEPRLLYSGSAIEGPAALQTESEAILRYSPEVAVDPDHLVQVOTVSYKF 119
DB 68 KSOEEEGNNRIITYQTAFHGLAQLTOEEAEERLEEEDGVAVIPETRYELHTRSTPF 127
QY 120 LGLDGFNGSVWKSFRGGQTIIGVLDTGWPESPFDGTGMPISIPRKWKGCQCEGSFS 179
DB 128 LGLERQESRVAERVTDDVVGVLDGTGWPESESNFTGMSVPFATWRGACETGKRF 187
QY 180 SSCNKRKLIGARFPIGRHVANSPSPNMPREYISARSTGCHGTASTVGSGSSVMAN 239
DB 188 KXCNKRKIVGARFYRYGEAAATGKIDE---ELEYKSPRDRDGHGTHTAATVAGSPVKGAN 244
QY 240 VLNGAGVARGMAGPAHAIVYKVCWPGNGCVSSDILAAIDVAIQDQVDVLSLGLGFP 299
DB 245 LFGFAGTAGMAQKARVAYKVCWVGCGCFSSDILSAVDQAVDGVQVLSISLGGGVSTY 304
QY 300 YDTIAIGTFRAMEGIVICRAGNNGPIESSVANTAPVWSTIGAGTLDRRPPAVRLAN 359
DB 305 SRSLSIATFGAMGMGVFFVSCSAGNGGPDPISLTNVSPWITTVGASTMDRDPFAVKIGT 364

```

```

QY 360 GKLLYGESLYPGKIKNAGREVVIYV---TGGDKGSEFCLRGSLPREETRGRKMWICDRG 416
DB 365 MRTFGVSLYKGRTVLPKNQYPLVYLGRLNASSPDP*SFCLDGLDRHRVAGKIVICDRG 424
QY 417 VNGRSEKEAVEKAGGVAMILANTEINOEDSIDVHLLPATLICYTESVLLKAYVATVK 476
DB 425 VTPRVQGVVVRAGGIGMVLNTATNGBELVADSHMLPAVAVGEGKGLIKQYAMTSKK 484
QY 477 PKARIIFGCTVIGRSRAPEVAQFSARGPSLANPILKPDMLAPGVNIIAAMPONLGP 536
DB 485 ATASLEILGTRIGIKPSFVVAFAFSRGNFELSLEILKPDLLAPGVNIIAAMPONL 544
QY 537 PYDSRRNVFTVMSGTSMSCPHVSIGITALIRSAVFNWSPAAIKSALMTTADLYDQ 596
DB 545 SSDPRVKFNLSGTSMSCPHVSIGVAALIKSRHDPWSPAAIKSALMTTAYVHDNM 604
QY 597 D--GNKAGVPAIGAAGHVNPCKAINPGLVNIQPDYITVILCTGLGTFRRSILAIT-HQ 653
DB 605 DASGAAPSSPYDHGAGHIDPLRATDPLVDIGPQEFELCTQDLSPLSQVKVFTK 664
QY 654 SCNGILRNKPGFSLNPSYIAVIFPKRGKTTMIT--RRVTNVGSENSIYSVNVKAPEGI 711
DB 665 TCKITLAKNPG-NLNYFAISALFPENTHVKAMTLRRVTNVGPHISSYKVSVPFRGASV 723
QY 712 IVPKRLVFRHVDQTLISY---RVWFVLYKKNRGKGVASFAQGQLTWNVSHNLMORVRSP 767
DB 724 TVQPKTLNFTSKHOKLSYTVTFRFRMKRPEFG-----LVWKSIT--THKVRSP 771
QY 768 ISVT 771
DB 772 VIIT 775

RESULT 12
Q39007
ID Q39007 PRELIMINARY; PRT; 746 AA.
AC Q39007;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Subtilisin-like protease (Fragment).
GN ARA12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
CX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=95375542; PubMed=7647567;
RA Ribeiro A., Akkermans A.D.L., Van Kammer A., Bieseling T.,
RA Pawlowski K.;
RT "A nodule-specific gene encoding a subtilisin-like protease is
RT expressed in early stages of actinorhizal nodule development.";
RL Plant Cell 7:785-794(1995).
RN [2]
RN SEQUENCE FROM OF 481-613 N.A.
RC STRAIN=cv. Columbia; TISSUE=Dry seed;
RA Raynal M., Grillet P., Laudie M., Meyer V., Cooke R., Delsey M.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DDJ databases.
DR EMBL; X85974; CAA59963.1; -.
DR EMBL; Z47391; CAA87460.1; -.
DR HSP; Q45670; IDBI.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00879; ODR_DC_2_2; 1.
DR PROSITE; PSS0840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.

```

```

KW Protease.
FT NON TER
SQ SEQUENCE 746 AA; 78212 MW; DC9A8783DF12C6DE CRC64;

Query Match
Best Local Similarity 40.8%; Score 1640.5; DB 10; Length 746;
Matches 359; Conservative 122; Mismatches 248; Indels 51; Gaps 19;

QY 6 FFLCIIIFLLFCSSSSSEILQKQYIVQLHPNSSETAKTFASKDFDHLSPFLQEAFLVGEVEEEE 65
DB 2 FFL-LCFLCHVSSSSSQGYIVHM-AKSTQ-PSFDLHNSNDSSURSISD--- 52
QY 66 EPSRLLYSGSAIEGFAQLTESEAEILRYSPVAVRDPDHLVQVQTYTSYKFLGLDG- 124
DB 53 --SAELLYYENALHGFSTELQCEADSLWTPGVISVLPEHRYELHTRTFLFLGLDEH 110
QY 125 ----FQNSGVMKSRFGQGTIIIVLDTGVWPSPSPDDTGMPISPRKWKGICQEGESRSS 180
DB 111 TADLFPFAGSYS-----DWVGVLDTGVWPSPSKSYSDGFGFIPSSWKGCCEAGTNTFA 164
QY 181 SSCNRKLI GARFTIRGHRVANSF-BESPNMPREYISARDSTGCHGTHTASTVGGSSVSMAN 239
DB 165 SLCNRKLI GARFARGYESTMTGPIDES-----KESRSRPRDDGCHGTHTSTAGSVEEGAS 220
QY 240 VLGNAGVARGMAPGAHIAVYKVCWFCNGCYSSDILAAIDVAIQDKVDVLSLSLGGFPPL 299
DB 221 LLGASGTARGMLHA--LAVYKVCWLGCGFSSDILAAIDKALADNVNVLNSLGGMSDY 278
QY 300 YDITIAIGTFRAMERGISVICAGNNGPIESSVANTAPWVSTIGAGTLDLDRFPVAVRIAN 359
DB 279 YRDGVAIGAFAMERGIIVSCSAGNAGPSSSLSNVAPWITTVGAGTLDLDRFPFALLIGN 338
QY 360 GKLLYESLPYKGIKNAGREVEIVVTGGDKGS--EFLCRLSPREIRGKXVLCDRGV 417
DB 339 GKNTFTGVSFLKGBALPD--KLLPFIYAGNASNATNGNLCMTGLIPEKYKGIWVMDRGI 396
QY 418 NGRSEKGEAKKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNATVKP 477
DB 397 NARVQGVVKAAGGVGMILANTAANGEELVADALHLLPATITVEEKAGDIIRHYVTTDNP 456
QY 478 KARIIFGGTVIGRSRAPEVAQFSARGPSILANPSILKPDMPAGVNIIAAPONLQPTGLP 537
DB 457 TASISILGTGVGKPSVVAASFSSRQPSNITNLIKPDLLIAPGVNIIAAGTGAAGPTGLA 516
QY 538 YDERRNFTVMSGTSMSCHVSGITAIIRSAVPNNSPAAIKSALMTTADLYDRQKAID 597
DB 517 SDSRRVEFNIIISGTSMSCHVSGLAALLKSVHPWSPAAIRSAIMTAYKTYKDGKPLID 576
QY 598 --GNKPAGVPAIGAGHVNPKALNPGIVYNIQPVYITLCTLGFTRSDILATHKVNVC 655
DB 577 IATGKPSPTPDHAGHVSPTTANPLIYDLTFTEDYLGLCALNTYSPQIRSVSRNYTC 636
QY 656 NGILRNKNGPS---LNVPSIAVIFKKGKTEMITRRTVNVGSPNSIYSVNVKA-PEGIKV 711
DB 637 D----PQSKSVADLNPSPFANVD-GAGAYKXTRVTSVGGAGT-YSVKVTSETTGKI 690
QY 712 INVPKLVFKHVDQTLISRYWVFLKKNRGKGVASFAQQLTWNASHNLMQVRSPISVT 771
DB 691 SVEPAVLNPKAEKKSYYTFTVDSKSPSG---SNSFGSIEMSDGKHV---VGSEFVALS 744

RESULT 13
P93204 PRELIMINARY; PRT; 766 AA.
AC P93204;
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Serine protease, SBT1.
GN SBT1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

```

OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RA Schaller A.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. VFW8;
RA Meichtry J., Amrhein N., Schaller A.;
RT "the gene family of subtilisin-like proteases in tomato.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; X98929; CAA67423.1; -.
DR EMBL; AJ006378; CAA06999.1; -.
DR HSP; O45670; IDBI
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease; Serine protease.
SQ SEQUENCE 766 AA; 81354 MW; 5D7ACCS5DCEFFESE7 CRC64;

Query Match
Best Local Similarity 40.1%; Score 1611; DB 10; Length 766;
Matches 339; Conservative 140; Mismatches 250; Indels 54; Gaps 17;

QY 8 LCIIIFLLFCSSSEIILQKQYIVQLHPNSSETAKTFASKDFDHLSPFLQEAFLVGEVEEEE 67
DB 13 MVVLFHVFDARQN--QKTYIIHMDKENMPA-----DFDDHTQWYDSSIKSVSK----- 60
QY 68 SSRLLYSYGAIEGFAQLTESEAEILRYSPVAVRDPDHLVQVQTYTSYKFLGLDGFGN 127
DB 61 SANMLYNSYVHGYSTQLTADAKALAOQPGILLVHEEVIVELHTRTSPTFLGLEGRS 120
QY 128 SGVMSKSRFCQGTIIIVLDTGVWPSPSPDDTGMPISPRKWGICQEGESFSSSCNRKL 187
DB 121 RSFPFQTEARSEVIIIVLDTGVWPSPSPDDTGMPISPRKWGICQEGESFSSSCNRKL 180
QY 188 IGRFFIRGRHVA-NSPESPNMPREYISARDSTGHTHTASTVGGSSVSMANVLNGAG 246
DB 181 IGRFFSQGYEAFAAIDET-----IESKPRDDEGHGTHATTAGSVTVGASLLGYATG 236
QY 247 VARGMAPGAHIAVYKVCWFCNGCYSSDILAAIDVAIQDKVDVLSLSLGGFPPIPYDITAI 306
DB 237 TARGMASHARVAAYKVCWTCGCFSSDILAGMDQAVDGVNVLSSLGTTISDVHRDIAI 296
QY 307 GTFRAMEGRISVICAAGNNGPIESSVANTAPWVSTIGAGTLDLDRFPVAVRLANGKLLYGE 366
DB 297 GAFSAASQGI FVSCSAGNGGFSSTLSNVAPWITTVGAGTMDREFPAYIIGNGKKLNGV 356
QY 367 SLVPGKGIKNAGREVEIVVTGGDKGS--EFLCRLSPREIRGKXVLCDRGVNRSSEK 424
DB 357 SLVSGKALPS--VMPLVYAGNVSSQSNGLCTSGSLIPEKVGAKIYVCDRGNNARQKG 414
QY 425 EAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNATVKPARIIFG 484
DB 415 LVVKDAGGIGMILANTDYGDELVAHLIPTAAVQGTAGNLIKQVIAASNPNATIAFG 474
QY 485 GTVIGRSRAPEVAQFSARGPSILANPSILKPDMPAGVNIIAAPONLQPTGLPYDERRVN 544
DB 475 GTKLGVQPSVVAASFSSRQPSNITPDVLPDLIAPGVNIIAGWTGKVGPTGLQEDTRNVG 534
QY 545 FTVMGTSMSCHVSGITAIIRSAVPNNSPAAIKSALMTTADLYDRQKAID-----GN 599
DB 535 FNIISGTSMSCHVSGLAALLKAAHPWSPAAIRSAIMTTSYSTYKNGKTIEDVATGMSS 594
QY 600 KPAGVPAIGAGHVNPKALNPGIVYNIQPVYITLCTLGFTRSDILATHKVNCSNGIL 659
DB 595 TP---FDYAGHVNPTAAVSFGLVYDLYTVDYINFLCALDYSPSMKVIKRDISCD--- 648

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 01:46:31 ; Search time 2735 Seconds
(without alignments)
6887.009 Million cell updates/sec

Title: US-09-806-767-2
Perfect score: 4018
Sequence: 1 MEKPPFLCIIIFLLFCSSS.....NSHNLQVRSPISVTLKTN 775

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2/1/USPRO.spool/US09806767/runat_02022004_085958_17911/app_query.fasta_1.967
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALCN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09806767@cgn_1_12132@runat_02022004_085958_17911 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_rpg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1282	31.9	832	28	BH480160
2	1092	27.2	696	28	BH660501
3	1079.5	26.9	1572	11	AY107161
C 4	1073	26.7	676	28	BH998381
C 5	1054	26.2	670	28	BH464077
C 6	1012	25.2	635	28	BH599290
C 7	979	24.4	640	28	BH442444
C 8	961.5	23.9	789	29	CC340521
9	923	23.0	627	28	BH597572
10	887.5	22.1	1795	11	AY108210
C 11	863	21.5	803	28	BH601308
12	820	20.4	750	28	AZ046831
13	775	19.3	767	28	BH793726
14	764	19.0	568	10	BF052773
C 15	749	18.6	527	28	BH714875
16	744	18.5	1006	11	AY105444
17	728.5	18.1	713	28	BH720365
18	715.5	17.8	719	13	BQ870650
19	700.5	17.4	844	14	CD434609
20	698	17.4	848	14	CD575777
21	696.5	17.3	849	14	CB642755
22	683.5	17.0	800	14	CB892659
23	680	16.9	859	14	CD573880
24	678	16.9	740	13	BQ867188
25	677	16.8	863	10	BF278122
26	676	16.8	873	10	BE216378
27	673.5	16.8	818	14	CD437637
28	669.5	16.7	697	29	BZ403171
C 29	665.5	16.6	710	13	BQ507537
30	662.5	16.5	768	29	BZ536866
31	661	16.5	709	13	BQ866922
C 32	656	16.3	853	28	BH649993
33	653	16.3	744	14	CB982404
C 34	651.5	16.2	878	29	BZ656313
35	646	16.1	738	13	BQ996378
36	636	15.8	770	10	BG598379
37	634	15.8	771	13	BU001251
38	632	15.7	771	13	BQ992191
39	630.5	15.7	926	10	BG445439
40	627.5	15.6	762	28	BH649984
41	627.5	15.6	923	29	BZ786320
C 42	625.5	15.6	766	28	BH725669
43	625.5	15.6	818	12	BI405859
C 44	624	15.5	864	28	BH648619
45	621.5	15.5	811	12	BG887274

ALIGNMENTS

RESULT 1
BH480160/c
LOCUS BH480160
DEFINITION BOHME48TF BOHM Brassica oleracea genomic clone BOHME48, genomic survey sequence.
ACCESSION BH480160
VERSION BH480160.1 GI:17688264
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.

```

1 (bases 1 to 832)
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOHME48TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
    source
        1..832
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone_lib="BOHM"
            /clone_name="BOHM48"
            /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
            genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      177 a   209 c   247 g   199 t
ORIGIN
Alignment Scores:
Pred. No.:      1,21e-123      Length:      832
Score:          1282.00        Matches:    244
Percent Similarity: 93.84%      Conservative: 15
Best Local Similarity: 86.41%    Mismatches:  16
Query Match:    31.91%         Indels:     1
DB:              28            Gaps:       0

US-09-806-767-2 (1-775) x BH480160 (1-832)
Qy      83  AlaAglNleuthrGluSerGluAlaGluLeuArgTyrSerProGluValValAla 102
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      832 TCAGCTCAGTTAACCGAATCAGAACCCAAAACACTCAAGACTTTACCTGAAGTTGTGCA 773
Qy     103 ValArgProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeu 122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     772 GTAAGACCTGACCATGTTCTCCAGTCCAAACCACTTACTCTTACCAAGTCTTAGAGACTC 713
Qy     123 AspGlyPheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleLeu 142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     712 AACGGTCCAGGACCCCTCCAGCGTATGGTCTAAATCCCGTCCGCCAAGGCACGATCATC 653
Qy     143 GlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetPro 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     652 GGGCGTTCTCGACACAGGAGCTCTGGCCCGAAGACCCCTAGCTTCGACGACACAGGGATGCCT 593
Qy     163 SerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSer 182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     592 TCAGTCCCAAGCAATGGAAGAGTGTTCACAGAGGAGAGAGCTTCACCTCTCTCAAC 593
Qy     183 CysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSer 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     532 TGTAAACAAGAACTAATCGGCGCTAGATTCTTCATCAGAGGCCACCGCGTAGCCAATCT 473
Qy     203 ProGluGluSerProAsnMetProArgGlnTyrIleSerAlaArgAspSerThrGlyHis 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     472 CTTTGGACTCATCCAAACATGCTCGCAATACATATCGGCACGAGACTCCACGGGGCAC 413
Qy     223 GlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGly 242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     412 GGGNCCACACACCCCTCAACTGCTGGCGGCTCTCTGTTCGACGGCGAGCGTTCGGT 353
Qy     243 AsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysVal 262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     352 AACGGACCCGGGTAGCCCGCGGATGGCGCCAGAGACTCAGTGGCGGTTCACAAATC 293
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY	263	CysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGln	282
Db	292	TGCTGGTTCAACGGCTGTGTACAGCTCAGACATCTCGCGCGATAGACGTAGCAATACAA	233
QY	283	AspIysValAspValLeuSerLeuGlyGlyPheProIleProLeuIleTrpValAsp	302
Db	232	GACAAAGTCAGCGTTCTATCCCTCTCCCTAGGCGGGTTTCCCTATCCCTTGTATGATGAC	173
QY	303	ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAla	322
Db	172	ACTATCGCCATTGGTACCTTCGCGGCACCGAATCTCCGTGTGCTCGCGGGT	113
QY	323	GlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIle	342
Db	112	GGCAACAACGGCGCGCTCGCTAGCTCGTGGCGGAACACAGCTCCTTGGGTCTCAACATC	53
QY	343	GlyAlaGlyThrLeuAspArgGlyPheProAlaValValArgLeu	357
Db	52	GGAGCAGGACGCTCGATAGAAAATTCCTCCGGGGGTGTGTAGTTA	7

RESULT 2	BH660501	696 bp	DNA	linear	GSS 19-FEB-2002
LOCUS	BOHWS711TF BO_2_3_KB Brassica oleracea genomic clone BOHWS71,				
DEFINITION	genomic survey sequence.				
ACCESSION	BH660501				
VERSION	BH660501.1	GI:18719169			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea				
ORGANISM	Brassica oleracea				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;				
TITLE	eurosid II; Brassicales; Brassicaceae; Brassica.				
JOURNAL	1 (bases 1 to 696)				
COMMENT	Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.				
	Whole genome shotgun sequencing of Brassica oleracea				
	Unpublished				
	Contact: Chris Town				
	TIGR				
	9712 Medical Center Drive, Rockville, MD 20850, USA.				
	Tel: 301-838-3523				
	Fax: 301-838-0208				
	Email: cdtown@tigr.org				
	DNA is from a doubled haploid provided by Tom Osborn.				
	Seq primer: TF				
	Class: sheared ends.				
FEATURES	Location/Qualifiers				
source	1..696				
	/organism="Brassica oleracea"				
	/mol_type="genomic DNA"				
	/strain="TO1000DH3"				
	/db_xref="taxon:3712"				
	/clone="BOHWS71"				
	/clone_lib="BO_2_3_KB"				
	/notes="vector: pHOSt1; Site_1: BstXI; 2-3 kb sheared				
	Genomic DNA inserted into pHOSt1 using BstXI linkers"				
BASE COUNT	163 a 166 c 220 g 147 t				
ORIGIN					

Alignment Scores:					
Pred. No.:	8,09e-104	Length:	696		
Score:	1092.00	Matches:	206		
Percent Similarity:	95.22%	Conservative:	13		
Best Local Similarity:	89.57%	Mismatches:	11		
Query Match:	27.18%	Indels:	0		
DB:	28	Gaps:	0		

US-09-806-767-2 (1-775) x BH660501 (1-696)

```

QY 249 AtgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys 268
Db 66 CGCGGATGCGCGAGAGTCACGTGGCGGTTCACAAAGTCTGCTGTTCAACCGCTGT 125
QY 269 TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeu 288
Db 126 TACAGCTCAGACATTCTGGCGGCGATAGAGTAGCGATACNAGACAAAGTCAGCTTCTA 185
QY 289 SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspThrIleAlaIleGlyThr 308
Db 186 TCGCTCTCCCTAGCGGTTCCCTATCCCTTTGTATGATGACATATCGCCATCGGTACC 245
QY 309 PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle 328
Db 246 TTCGCGGCCACGGAACATCTCGTTGTCTGCGCGGTGGCAACACGCGCGCTC 305
QY 329 GluSerSerValAlaAsnThrAlaProTyrValSerThrIleGlyAlaGlyThrLeuAsp 348
Db 306 GCTAGTCTCGGTGGCAACACAGCTCTTGGTCTCAACCATCGGACGAGAACGCTCGAT 365
QY 349 ArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeu 368
Db 366 AGAAATTCCTCGCGGTTGTAGTTAGCAACGGGAGCTACTCTACGAGAGTCGTG 425
QY 369 TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGly 388
Db 426 TACCCAGGAAAGGATAAAGAGGCTGAGAGAGTTGGAGATTCTACCTGGCGGGA 485
QY 389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluLulileArgGly 408
Db 486 GGAGACAAAGGAGGAGATTTCCTTGAGAGGCTCGCTTCCGAGAGAGTGTCCAAAGC 545
QY 409 LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys 428
Db 546 AGATGGTGATATGCCAGCTGGAGTCACCGGAGATCAGAGAGGACAGCGGTTAAA 605
QY 429 GluAlaGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSer 448
Db 606 GAAGCTGGCGCTCGGATGATCTTGGCTTAACACCGAGATTAAACCAAGAGGAGATTCT 665
QY 449 IleAspValHisLeuLeuProAlaThrLeu 458
Db 666 GTAGACGTTTCATTGTACACGAGCTAGCTTG 695

RESULT 3
AY107161 1572 bp mRNA linear HTC 16-OCT-2002
LOCUS AY107161
DEFINITION Zea mays PC0101731 mRNA sequence.
ACCESSION AY107161
VERSION AY107161.1 GI:21210239
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1572)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 1572)
Coe, E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:

```

```

www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source            1..1572
     /organism="Zea mays"
     /mol_type="mRNA"
     /db_xref="MaizeDB:635852"
     /db_xref="taxon:4577"
     /clone_lib="Maize Mapping Project/DuPont Consensus
     Library"
     /note="this sequence is part of a project of EST
     assemblies resulting from the application of public
     contigs to seed DuPont contigs; this resource was
     assembled by DuPont as part of a collaboration for the
     overgo addressing of BACs in conjunction with the Maize
     Mapping Project"
BASE COUNT          297 a 542 c 465 g 268 t
ORIGIN
Alignment Scores:
Pred. No.:          5.66e-102          Length:          1572
Score:              1079.50           Matches:         230
Percent Similarity: 61.72%           Conservative:    86
Best Local Similarity: 44.92%        Mismatches:     177
Query Match:        26.87%           Indels:         19
DB:                 11                Gaps:           10
US-09-806-767-2 (1-775) x AY107161 (1-1572)

```

```

QY 268 CysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspVal 287
Db 40 TGCTTCAAGTCCGACATCTCTGGCGCGCTGCGCGCGCTGCGCGCGCGCGCGCGCTG 99
QY 288 LeuSerLeuSerLeuGlyIlePheProIleProLeuTyrAspThrIleAlaIleGly 307
Db 100 CTCTCCATCTCGCTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 159
QY 308 ThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyPro 327
Db 160 TCGTTCGTGTCATGAGATGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 219
QY 328 IleGluSerSerValAlaAsnThrAlaProTyrValSerThrIleGlyAlaGlyThrLeu 347
Db 220 GACCCCAATAAGCTTCACCAACCTGTCTGCGCTGGATACCAACGCTGGCGCGCGCGCGCG 279
QY 348 AspArgArgPheProAlaValAlaArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSer 367
Db 280 GACCGGAGCTTCCCGGCGACGCTGCGCAATGGCGCAACCTTCACCGCGGTTTCG 339
QY 368 LeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThr 387
Db 340 CTCTACAAGCGCGCGAGGTCTTCTGCTCAAGGAGCAGTACCCGTTAGTCTACATGGCG 399
QY 388 GlyGlyAspLys-----GlySerGluPheCysLeuArgGlySerLeuProArgGlu 404
Db 400 GGCAACTCGAGCATCTCCGACCCAGCTCCCTGCTGCTGCTGCGGAGGACACTCCAGCCCGAC 459
QY 405 GluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGly 424
Db 460 GAGTGTGCTGGAAGATTGTGATCTGCGACGCGGCGCATTAGTCTCGGTCGCGTCGAGAGG 519
QY 425 GluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln 444
Db 520 CAGTTGTCAAGAACGCGCGGTGCGCGGCGATGATCTCGCCCAACACCGCGCAACCGCG 579
QY 445 GluGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSer 464
Db 580 GAGAGGTTGTCGCGGACAGCCACCTCTCCAGCAGTGGCGCGTGGCGAGTCTGAAGCG 639
QY 465 ValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGly 484
Db 640 ATTCCGCCCAAGAGTAGTACAGCAAAACCGCCCGCAACCAACCGCGCACGCTCAGCTTCGAC 699
QY 485 GlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyPro 504

```

```

700 GGGACGAAGCTCGGATCCGCCATCGCCAGTCGTAGCGCGTCTTCGTCGCCGGGACCA 759
505 SerLeuAlaasnProSerIleLeuLysProAspMetIleAlaProGlyValasnIle 524
760 AATCTCTGCTACCTCGAGATCCTCAAGCCGGAGTTCATCGCCCGCGGTGAACATCTG 819
525 AlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsn 544
820 CGGCATGGAGCGCGACCGCCGCTCGAGCTTGTCCAGCAGCGCGCGGTGCGC 879
545 PheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeu 564
880 TTCAACATCTGTCGGGAGCTCATGTCGTGCCGACGTCGCGCGCTGCGCTGCGCTG 939
565 IleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThr 584
940 ATCAAGCCACCGACCGGACCTGAGCCCGCCGGAAGATCAAGTCGCGCTGATGACCC 999
585 AlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAla 602
1000 GCGTACGTCACACACACAGTACCGGTGCTCGAAGAGCGGCCACCGCGAAGCGCTCC 1059
603 GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeu 622
1060 ACGCGTTCGATCAGCGAGCTGGGCATACACCCGCTGCTGTCCTCAACCTCGCGCTG 1119
623 ValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArg 642
1120 GTCTACGACATCGCCAGGACGACTACCTGGAGTTCCTCTGCTGGAGAACCTGACGCG 1179
643 SerAspIleLeuAlaIleThrHisLysAsnValSer-----CysAsnGlyIleLeuArg 660
1180 TTGACGCTCAGGTCCTTCACC---AAGACTCGAGCAAGACATGCAAGCACCTTC--- 1233
661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
1234 AGCTCGCGCGGT---GACTTGATTATCCGGCCATCTCCGCGTCTTCGCGAGCAGCCA 1290
681 ThrThrGluMetIleThrArgArg---ValThrAsnValGlySerProAsnSerIleTyr 699
1291 TCTGCTCGCTGACGCTGCGTGCACCGTCGACGAGCGTGCAGAACGTCGCGCGCGCTTC 1350
700 SerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuVal 719
1351 CATGTCGAAGTTACAGAGTTCAAGGCGCAGACATGTCGTGCAACCAAGACACCTGCG 1410
720 PheLysHisValAspGlnThrLeuSerTyrArgValTrpPheValLeuLysLysAsn 739
1411 TTCAAGTTTCGAACCAAGAGCTAACCTACAGGTGACGATGACACCAAGCGTGCAC 1470
740 ArgGlyLysValAlaSerPheAlaGlnGlyLysLeuThrTrpValAsnSerHisAsn 759
1471 -----AAGACACCGGAGTTC-----GGAGCGCTGCTTTGAGCAGCGCGTCCAC 1515
760 LeuMetGlnArgValArgSerProIleSerValThr 771
1516 ATC-----GTCCGGAGCCCGCTTGTCTCCACA 1542

RESULT 4
LOCUS BH998381/c
DEFINITION oeh58e11.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION BH998381
VERSION BH998381.1
KEYWORDS GI:23544689
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 676)

```

```

AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T., Nash
W., Rabinowicz, P.D. and Willson, R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Richard K. Willson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Plate: oeh58 row: e column: 11
Seq primer: -21UPpOT forward
Class: shotgun
High quality sequence start: 14
High quality sequence stop: 551.
Location/Qualifiers
1..676
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
BASE COUNT 138 a 177 c 208 g 153 t
ORIGIN
Alignment Scores:
Pred. No.: 7,64e-102 Length: 676
Score: 1073.00 Matches: 200
Percent Similarity: 94.20% Conservative: 11
Best Local Similarity: 89.29% Mismatches: 13
Query Match: 26.70% Indels: 0
DB: 28 Gaps: 0
US-09-806-767-2 (1-775) x BH998381 (1-676)
QY 118 LysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysArgPheGly 137
DB 675 AAGTTCTTAGACATCAACGGTCCAGGACCTCCAGCGTATGCTCTAAATCCGCTCCGCG 616
QY 138 GlnGlyThrIleLeGlyValLeuAspThrGlyValTrpProGluSerProSerPheAsp 157
DB 615 CAAGCGCAGATCATCGCGTTCGACACAGGAGTCTGGCCCGGAAAGCCCTTAGCTTCGAC 556
QY 158 AspThrGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGlyGluSer 177
DB 555 GACACAGGATGCTCTCAGTCCCAAGCAATGGNAAGTGTTCGCCAAGAGGAGAGCAGC 496
QY 178 PheSerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHis 197
DB 495 TTCACCTTCCTCAAACTGTACACAGAACTAATCGCGCTAGATTCTTCATCAGAGGCGAC 436
QY 198 ArgValAlaAsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArg 217
DB 435 CGCGTAGCCAACTCTCCTTTGGACTCCAAACATGCTCCGGAATACATATCGGCACGA 376
QY 218 AspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMet 237
DB 375 GACTCCACGGGACGAGGACCCACACCGCTCAACTGCTGGCGCTCCTCTGTTTCACG 316
QY 238 AlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIle 257
DB 315 CCGAGCGTTCGCTAAGCGAGCGCGGTAGCCCGCGGGATGGCCGAGAGCTCAGCTG 256
QY 258 AlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaIle 277
DB 255 CGCGTTTACAAAGTCTCTGTTCAAGCGCTGTTTACAGCTCAGACATCTCTGGCGCGATA 196
QY 278 AspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIle 297

```

```

FEATURES
source
1..676
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

```

```

Db      195 GAGGTAGCATACAGACAAAAGTCGACGTTCTATCCCTCCTAGGCGGTTTCCTATC 136
Qy      298 ProLeuTyAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSer 317
Db      135 CCTTTGATGATGACACTATGCCATTGGTACTCTTCGCGCCACGGAACATGGAATCTCC 76
Qy      318 ValIleCyAlaAlaGlyAsnGlyProIleGluSerSerValAlaAsnThrAlaPro 337
Db      75 GTTGTCTGCGCGGTGGCAACACGCGCGCTCGCTAGCTCGGTGGCAACACAGCTCCT 16
Qy      338 TrpValSerThr 341
Db      15 TGGGATTCACC 4

RESULT 5
BH464077/c
LOCUS      670 bp      DNA      linear      GSS 13-DEC-2001
DEFINITION BOGEF01TF BOGE Brassica oleracea genomic clone BOGEF01, genomic
survey sequence.
ACCESSION      BH464077
VERSION      BH464077.1 GI:17659406
KEYWORDS      GSS
SOURCE      Brassica oleracea
ORGANISM      Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 670)
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
1..670
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGEF01"
/clone_lib="BOGE"
/note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOSt using BstXI linkers"
BASE COUNT      137 a 188 c 186 g 159 t
ORIGIN

Alignment Scores:
Pred. No.:      7,44e-100      Length:      670
Score:      1054.00      Matches:      197
Percent Similarity:      94.14%      Conservative:      12
Best Local Similarity:      88.74%      Mismatches:      13
Query Match:      26.23%      Indels:      0
DB:      28      Gaps:      0

US-09-806-767-2 (1-775) x BH464077 (1-670)

Qy      158 AspThrGlyMetProSerIleProArgLysTrpLysGlyIleCyGlnGluGlyGluSer 177
Db      670 GACACAGGATGTTTCATCCACCAATGGAAGGTGTTTGCAGAGGAGAGAGC 611
Qy      178 PheSerSerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHis 197
Db      610 TTCACTCTTCAACTGTTCAACAGAACTAATTCGCGGTAGATTCTTCATCAGAGGCCAC 551
Qy      198 ArgValAlaAsnSerProGluGluSerProAsnMetProArgGluTrpIleSerAlaArg 217

```

```

Db      550 CGGTAGCAACACTCTCTTTGGACTCACCAAAACATCGCTCGGAATACATATCGGCACGA 491
Qy      218 AspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMet 237
Db      490 GACTCCACGGGGCACGGGACCCACACCGCTCAACTGCTGGCGGCTCTCTGTTCGACG 431
Qy      238 AlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIle 257
Db      430 GCGAGCGCTTCGTGTAACGGAGCGGGGTAGCCCGGGATGCGCCGAGGAGCTCAGTG 371
Qy      258 AlaValTrpLysValCysTrpPheAsnGlyCysTrpSerSerAspIleLeuAlaIle 277
Db      370 GCGTTTTACAAAGTCTGTGTTCAACGGCTGTTCAGCTCAGACATTCTGGCGGCGATA 311
Qy      278 AspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIle 297
Db      310 GAGGTAGCATACAGACAAAGTCGAGCTTCATCTCTCCCTAGGCGGTTTCCCTATC 251
Qy      298 ProLeuTyAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSer 317
Db      250 CCTTTGATGATGACACTATCGCAATTGGTACCTTCGCGCCACGGAACATGGAATCTCC 191
Qy      318 ValIleCyAlaAlaGlyAsnGlyProIleGluSerSerValAlaAsnThrAlaPro 337
Db      190 GTTGTCTGCGCGGTGGCAACACGCGCGCTCGCTAGCTCGGTGGCGAACACAGCTCCT 131
Qy      338 TrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValAlaArgLeu 357
Db      130 TGGGTCTCAACCATCGAGCAGAACGCTCGATAGAAATTCGCGGGGTGTGTAGTTA 71
Qy      358 AlaAsnGlyLysLeuLeuTyGlyGluSerLeuTyProGlyLysGlyIleLysAsnAla 377
Db      70 GCCACGGGAAGCTACTCTACGAGAGTCTGTACCCAGGAAGGATAGAGAGCGGA 11
Qy      378 GlyArg 379
Db      10 GAGAGA 5

RESULT 6
BH59290/c
LOCUS      635 bp      DNA      linear      GSS 15-DEC-2001
DEFINITION BOGWR58TR BOGW Brassica oleracea genomic clone BOGWR58, genomic
survey sequence.
ACCESSION      BH59290
VERSION      BH59290.1 GI:17851736
KEYWORDS      GSS.
SOURCE      Brassica oleracea
ORGANISM      Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 635)
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other_GSSs: BOGWR58TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1..635
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGWR58"
/clone_lib="BOGW"

FEATURES
source

```

/note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS01 using BstXI linkers"

BASE COUNT ORIGIN

127 a 198 c 162 g 148 t

Alignment Scores:

Pred. No.: 1.75e-95 Length: 635
Score: 1012.00 Matches: 190
Percent Similarity: 94.79% Conservative: 10
Best Local Similarity: 90.03% Mismatches: 11
Query Match: 25.19% Indels: 0
DB: 28 Gaps: 0

US-09-806-767-2 (1-775) x BH599290 (1-635)

QY 213 TTTTCTGTTTTCAGCGGAGCTTCTCGGTAAACGAGCGGGGTAGCCCGGGATGGCG 514
Db 633 TACATATCGCAGAGCTCCACGGCCACCGGACCCACCGCTCAACTGCTGGCGGC 574
QY 233 SerSerValSerMetAlaAsnValLeuGlyAanGlyAlaGlyValAlaArgGlyMetAla 252
Db 573 TCTCTGTTTTCAGCGGAGCTTCTCGGTAAACGAGCGGGGTAGCCCGGGATGGCG 514
QY 253 ProGlyValHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTrpSerSerAsp 272
Db 513 CCAGGAGCTCAGTGGCGGCTTTTCAAGTCTCTGTTCAACGGCTGTTCAGCTCAGAC 454
QY 273 IleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeu 292
Db 453 ATTCTGGCGGATAGACGATACGATACAAAGTCAAGTCAAGTCTTCTATCCTCTCCCTA 394
QY 293 GlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMet 312
Db 393 GCGGTTTCCCTATCCTTTGATGATGACACTATCGCCATTGTTACCTTCGCGCCACG 334
QY 313 GluArgGlyIleSerValIleCysAlaAlaGlyValAsnAsnGlyProIleGluSerVal 332
Db 333 GAACATGGAATCTCCCTGTTGCGCGGTGCAACACCGGCCCTCGCTAGCTCGGTG 274
QY 333 AlaAsnThrAlaProIlePheSerThrIleGlyAlaGlyThrLeuAspArgPhePro 352
Db 273 GCGAACACAGCTCTTGGTCTCAACCATCGGAGCAGGACGCTCGATAGAAATTCCTCG 214
QY 353 AlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLys 372
Db 213 GGGGTTGTTAGTTAGTCAACGGAAGCTACTCTACGAGAGTCTGTTCACCGAGGAAA 154
QY 373 GlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyLysPlyGly 392
Db 153 GGGATAAAGAGCTCAGAGAGAGTGGAGATTGTCTACGTGCGCGGAGGAGACAAAGGA 94
QY 393 SerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIle 412
Db 93 AGCGAGTTTGTGTTGAGAGCTCGCTTCGAGAGAGAGTGTCCAAAGGCAAGATGTGTATA 34
QY 413 CysAspArgGlyValAsnGlyArgSerGlyLys 423
Db 33 TGGACCGTGGAGTCAACGGGAGATCAGAGAAA 1

RESULT 7

BH442444/c 640 bp DNA linear GSS 12-DEC-2001
LOCUS BOHGT85TR BOHG Brassica oleracea genomic clone BOHGT85, genomic
DEFINITION survey sequence.
ACCESSION BH442444
VERSION BH442444.1 GI:17628158
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 640)

AUTHORS TITLE JOURNAL COMMENT

Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of *Brassica oleracea*
Unpublished
Other_GSSs: BOHGT85TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES source

1..640
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOHGT85"
/clone_lib="BOHG"
/note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS01 using BstXI linkers"

BASE COUNT 149 a 211 c 126 g 154 t
ORIGIN

Alignment Scores:

Pred. No.: 5.16e-92 Length: 640
Score: 979.00 Matches: 184
Percent Similarity: 96.15% Conservative: 16
Best Local Similarity: 88.46% Mismatches: 8
Query Match: 24.37% Indels: 0
DB: 28 Gaps: 0

US-09-806-767-2 (1-775) x BH442444 (1-640)

QY 396 CysLeuArgGlySerLeuProArgGluIleArgGlyLysMetValIleCysAspArg 415
Db 626 TGTCTGAGAGCTCGTCCCGAGAGAGTGTCCAAAGCAAGATGGTGTATGCGACCTG 567
QY 416 GlyValAsnGlyArgSerGlyLysGlyAlaValLysGluAlaGlyValAlaMet 435
Db 566 GGAGTCACCGGAGATCAGAAAGACAGCGGTTTAAAGAGCTGGCGCGCTCGCGATG 507
QY 436 IleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuPro 455
Db 506 ATCTTGCTACCAACCGAGATTACCAAGAGGAGGATCTGTAGACCTTCATTGTATACCA 447
QY 456 AlaThrIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrVal 475
Db 446 GCTACGTTGATTGTTTACGAAGAGTCTGTGTTTGAAGGTTTACGTGAGGACACGGTA 387
QY 476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db 386 AGACCGAAGCTAGTTAATTTTCGCGGAGCGGTGATTTGGAGGTCGAGAGGCGCTGAG 327
QY 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
Db 326 GTGCGCAGTTCTCAGCGAGGACCGAGTTTGGCTTACCCCTTCGGTGTCTTAAACCGGAT 267
QY 516 MetIleAlaProGlyValAsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGly 535
Db 266 TTGATCGCGCAGGTGTCAATATTATAGTCTGTTCCTCAGAACTTCGAGCGCGGG 207
QY 536 LeuProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCys 555
Db 206 CTTCCTTATGATTCGAGGAGGTTAACTTCACGTGATGTCGGGAGCTTCGATGTTGT 147
QY 556 ProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAla 575
Db 146 CCGCACGTGAGTGGATCACCCTCTCATCGGCTCTCGGTATCCGAGCTGGTCTCCAGCT 87
QY 576 AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIle 595

Qy	471	ValAsnAlaThrValLysProIysAlaArgIleIlePheGlyThrValIleGlyArg	490
Db	547	ATCAGTCTGACGCCACGGCCGGTGGGAGATCGTATTCGGCGGCACCGCGATCGGACGA	488
Qy	491	SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer	510
Db	487	GCAGCGCCCGCGCGGTGGCGGTCTCTCGCGCGCGCGCGGAGCTGACCAACCCGTCG	428
Qy	511	IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaIleAlaAsn	530
Db	427	GTGCTGAAGCCCGACGTCGTGCGCCCGGGGTGAACATCATCATCGCGCGTGGCCGGGAC	368
Qy	531	LeuGlyProThrGlyLeuProTyrAspSerArgValAsnPheThrValMetSerGly	550
Db	367	CTGGGCGCGTTCGGGGGTGGAGAGCGACGCCCGCGGTCCCAACTTCAACGTCCTCGGG	308
Qy	551	ThrSerMetSerCysProHisValSerGlyIleThrAlaIleLeuIleArgSerAlaTy-Pro	570
Db	307	ACATCATGATGCGCGGCTCTATGATGATGGCATCGCGCGCTGATCCGTCGGCGCACCCG	248
Qy	571	AsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArg	590
Db	247	TCCTGGAGCCCGGATGTCGGTCCGGTATCATGACACCGGTGACATAATACGACCGG	188
Qy	591	GlnGlyLysAlaIleLysAspGlyAsnLysProAlaGly-----ValPheAlaIle	607
Db	187	CAGGGGAAGCGCATCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATG	128
Qy	608	GlyValAclyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGln	627
Db	127	GCGCGCGCGCACGTGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	68
Qy	628	ProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIle	645
Db	67	CCCGCGCACTACTGACGACCTGTGCACGCTCGGGTACACCCACATCGGAGATC	14
RESULT 9	BH597572	627 bp DNA linear	GSS 15-DEC-2001
LOCUS	BOHON29TF BOHO Brassica oleracea genomic clone BOHON29, genomic		
DEFINITION	survey sequence.		
ACCESSION	BH597572		
VERSION	BH597572.1	GI:17850024	
KEYWORDS	GSS.		
SOURCE	Brassica oleracea		
ORGANISM	Brassica oleracea		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids		
TITLE	/ eucrotids II; Brassicales; Brassicaceae; Brassica.		
JOURNAL	1 (bases 1 to 627)		
COMMENT	Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.		
	Whole genome shotgun sequencing of Brassica oleracea		
	Unpublished		
	Other_GSSs: BOHON29TR		
	Contact: Chris Town		
	TIGR		
	9712 Medical Center Drive, Rockville, MD 20850, USA.		
	Tel: 301-838-3523		
	Fax: 301-838-0208		
	Email: cdtown@tigr.org		
	DNA is from a doubled haploid provided by Tom Osborn.		
	Seq primer: TF		
	Class: sheared ends.		
FEATURES	Location/Qualifiers		
source	1..627		
	/organism="Brassica oleracea"		
	/mol_type="genomic DNA"		
	/strain="TO100DH3"		
	/db_xref="taxon:3712"		
	/clone="BOHON29"		
	/clone_lib="BOHO"		
	/notes="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared		
	genomic DNA inserted into pHOSt using BstXI linkers"		

Overgo Probes
Unpublished (2002)
2 (bases 1 to 1795)
Coe, E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES
Location/Qualifiers
1..1795
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:635508"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT 332 a 587 c 603 g 273 t
ORIGIN

Alignment Scores:
Pred. No.: 9,78e-82 Length: 1795
Score: 887.50 Matches: 199
Percent Similarity: 58.19% Conservative: 78
Best Local Similarity: 41.81% Mismatches: 174
Query Match: 22.09% Indels: 25
DB: 11 Gaps: 12

US-09-806-767-2 (1-775) x AY108210 (1-1795)

QY 302 AspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAla 321
DB 48 GACCTCATCGCCATCGCCACCGTTCAGGSCCATGAGACCGGATCTTCGTGACGGCCGCG 107
QY 322 AlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrrpValSerThr 341
DB 108 GGGGCGCATGACGGCCCGCGGGGCTCCATCACCACCGCGGCCCATGATGCTCAGC 167
QY 342 IleGlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGlyLys 361
DB 168 GTGCGGGCGGGCACCATGGACCGCGCATGACACCGGTCAGGCTCGCGACGGCAG 227
QY 362 LeuLeuTyrrGlyLysSerIleuTyrrProGlyLysGlyIleLysAsnAlaGlyArgGluVal 381
DB 228 GTGTTGACGCGGAGTCTGCTGTCAGCCCGGAGAAC---AACACGGGGGGCGCCCGCTG 284
QY 382 GluValIleTyrrVal---ThrGlyLysAspLysGlySerGluPheCysLeuArgGlySer 400
DB 285 CGCTCGTCTTCCCGGGCGGCAACCGCGACCCCGGAGGCCCGCGACTGC-----AGCAGC 338
QY 401 LeuProArgGluGluIleArgGlyLysMetValIleCysAsp---ArgGlyValAsnGly 419
DB 339 CTGGTGGAGCGCGAGGTGAGGGCAAGGTGTGTGTGTCGAGAGCCGCTCCATCAGCGAG 398
QY 420 ArgSerGluLysGlyGluAlaValLysGluAlaGlyLysValAlaMetIleLeuAlaAsn 439
DB 399 CACGTGACGAGCGGACGATGTGTCCCGTACCGCGCGCGCGGATGATCTCTCATGAAC 458
QY 440 ThrGluIleAsnGlnGluLysAspSerIleAspValHisLeuLeuProAlaThrLeuIle 459
DB 459 AGCCGGCGGAGGGGTTCACACCTTCGCCGACGCGGACGTCCTGCGGGCTGCCACGTG 518

BASE COUNT 161 a 186 c 146 g 134 t
ORIGIN

Alignment Scores:
Pred. No.: 3,77e-86 Length: 627
Score: 923.00 Matches: 177
Percent Similarity: 92.34% Conservative: 16
Best Local Similarity: 84.69% Mismatches: 15
Query Match: 22.97% Indels: 1
DB: 28 Gaps: 0

US-09-806-767-2 (1-775) x BH597572 (1-627)

QY 50 LeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluGluGluProSerSer 69
DB 1 CTCTCTTTCTTCAAGAAGCTGTCTAGGTATTGAAGAAGAAACGAAGACCTTCTACA 60
QY 70 ArgLeuLeuTyrrSerTyrrGlySerAlaIleGluGlyPheAlaAlaGluLeuThrGluSer 89
DB 61 AGGATTCTCTACTCTTACGCCCTCTGCCATTGAGGATCTCAGCTCACTTAACCGAATCA 120
QY 90 GluAlaGluLeuArgTyrrSerProGluValValAlaValArgProAspHisValLeu 109
DB 121 GAGCCMAACACTCAAGACTTACCTGAGTTGTTCAGTAGACCTGACCACTGTCTC 180
QY 110 GlnValGlnThrTyrrSerTyrrTyrrPheLeuGlyLeuAspGlyPheGlyAsnSerGly 129
DB 181 CAAGTCMAACCACTTACTCTTACAACTCTTAGGACTCAACGGTCCAGGACCCCTCCAGC 240
QY 130 ValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal 149
DB 241 GTATGTTCTAAATCCCGGTCCGCGCAAGCACGATCATCGGCGTCTTCGACACAGGAGTC 300
QY 150 TrpProGluSerProSerPheAspThrGlyMetProSerIleProArgLysTrpLys 169
DB 301 TGCCCCGAAGCCCTAGCTTCGACGACACAGGATGCCTTCAGTCCCAAGCAATGGAAA 360
QY 170 GlyIleCysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGly 189
DB 361 GGTGTTTGCCAAAGAGGAGAGCTTCACTTCCTCCACTGTAACAGAAACTAATCGGC 420
QY 190 AlaArgPhePheIleArgGlyHis-ArgValAlaAsnSerProGluGluSerProAsnMe 209
DB 421 GCTAGATTCTTCATCAGAGGCCACCGCGGTAGCCAACTCTCCTTTGGACTCACCACAAAT 480
QY 209 TrpArgGluTyrrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerTh 229
DB 481 GCCTCGGGAATACATATCGGACGAGACTTCCAGGGGACCGGACCCACACCGCTCAAC 540
QY 229 rValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyValAlaGlyValAlaAr 249
DB 541 TGCTGGGGCTCTCTGTTCGACGGCGAGCGTCTTCGTTACGAGGCCCGGGTACGCCG 600
QY 249 gGlyMetAlaProGlyAlaHisIle 257
DB 601 CGGGATGGCGGAGGAGCTCACGTG 625

RESULT 10
AY108210
LOCUS
DEFINITION Zea mays PC0096605 mRNA sequence.
ACCESSION AY108210
VERSION AY108210.1 GI:21211288
KEYWORDS HTC
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1795)
Hainey, C.F., Dolan, M., Mao, G.H., Vogel, J.M., Whitesitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of

REFERENCE
AUTHORS
TITLE

```

QY 460 GlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAla 479
Db 519 AGCTACCGCGCGGGTCCAGATCGCGGGGTATCAAGTCAGCGCCAGGCGCCAGCGG 578
QY 480 ArgIlePheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPhe 499
Db 579 ACATACAGTTCGGGGCACGGTATGGTCTGCTCCCGCGCGCGTCCGTCGCTTCCTTC 638
QY 500 SerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaPro 519
Db 639 TCGTCGCGCGCGCGAAGCGAGCGCGGCATCTCTGAAGCCGACATCACCGGGGCC 698
QY 520 GlyValAsnIleLeuAlaLafp---ProGlnAsnLeuGlyProThrGlyLeuProTyr 538
Db 699 GGAATGAATCATCTGCGCGCGTGGCGCCGAGCGAGCATGACCGGAGTTCGCG----- 752
QY 539 AspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisVal 558
Db 753 GACGAGTGAAGCTGCGCTTCTCATGGAGTCGGGACGCTCATGTCGAGCGCGCACCTG 812
QY 559 SerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaAlaIleLys 578
Db 813 AGCGGATCGCGGCATCATCAAGACCTGCACCGAGTGTGTCCTCCCGCGGATCAAG 872
QY 579 SerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyAlaIleLysAspGly 598
Db 873 TCGGCGATCATGACGTCGTCGCGCACCGCGGACCGCGGCGTGCAGATCAAGACGAG 932
QY 599 Asn---LysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAla 617
Db 933 CAGTACAGCGCGCGAGCTTACAGCATGGCGCGCGTACGTGAACCGCTCCCGCGCC 992
QY 618 IleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThr 637
Db 993 GTGACCGCGCGTGTGTACAGCTCGCGCGCGGAGTACATCGCTACCTCTCGCGGG 1052
QY 638 LeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGly 657
Db 1053 CTGGGATCGGGACGCGCGTGAAGAGATAACTGGCGCGCGCTGCGCTGC---GCC 1109
QY 658 IleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLys 677
Db 1110 AAGCTGAAGGCCATCACCGAGCGGAGCTCAACTACCGCTCGCTGTGTGAAGTGCTG 1169
QY 678 ArgGlyValThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSer 697
Db 1170 TCCATPCGATCACC-----GTCCCGCGCACCGTGAACCAAGTCCGGAAGCCAACTCG 1223
QY 698 IleTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArg 717
Db 1224 GTGTACAAGCGCGTGTGTGACATGCCAGCGCGGTGTCGCTGCGTGGCGCGCCCGGTG 1283
QY 718 LeuValPheLysHisValAspGlnThrLeuSerTyrArgVal-----TrpPheVal 734
Db 1284 CTGCGCTTCGCCAGCGCGAAGAGAGAGAGCTTACCGTACCGTGCCTGCGG----- 1337
QY 735 LeuLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGlyClnLeuThrTrp 754
Db 1338 -----AACGGCGCGCGCGCGTGCAGCGCGCGCGAGGGAACCTGAAGTGG 1382
QY 755 ValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
Db 1383 GTGTGAGGAGCAGCATC-----GTCCGAGCGCGCATCGTTATA 1421

```

RESULT 11

```

BH601308/c 803 bp DNA linear GSS 15-DEC-2001
LOCUS BOHSA897F BOHS Brassica oleracea genomic clone BOHSA89, genomic
DEFINITION survey sequence.
ACCESSION BH601308
VERSION BH601308.1 GI:17853754
KEYWORDS GSS.
SOURCE Brassica oleracea

```

```

ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 803)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other GSSs: BOHSA89TR
Contact: Chris Town
TIGR Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: Sheared ends.
FEATURES
source
1..803
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHSA89"
/clone_lib="BOHS"
/vector="Vector: PHOS1; site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT 245 a 223 c 121 g 214 t
ORIGIN
Alignment Scores:
Pred. No.: 1,09e-79 Length: 803
Score: 863.00 Matches: 167
Percent Similarity: 91.13% Conservative: 18
Best Local Similarity: 82.27% Mismatches: 16
Query Match: 21.48% Indels: 2
DB: 2 Gaps: 2
US-09-806-767-2 (1-775) x BH601308 (1-803)
QY 575 AlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAla 594
Db 803 GCTGCTATCAGATCAGCGATGATGACGCGCTGATTGTATGATCGAGAGGAAAGAG 744
QY 595 IleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db 743 ATTAGGACGCGGACAAACCGCGGAGTGTGTCTATTGGACGAGCGCATGTGAATCCG 694
QY 615 GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr 634
Db 683 GTTAAGCGGATTAACCGGGGTGGTTTACAACATCAACCGGTTGATTACATTGCTTAC 624
QY 635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer 654
Db 623 TTATGCACTATGGTTTACTAGATCGATATTTTAGCATCTACTCACAGAACGTTAGC 564
QY 655 CysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaVal 674
Db 563 TCGGTCCGCTACTGAGGAAGACCCCGGTTTGTAGTCTTAATATCCGCTATTTCGGTT 504
QY 675 IlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThrAsnValGlySer 694
Db 503 ATTTTAGCGTGGGAGACTAAGGAGATGATCAGGAGCGGTGTGACTAACGTTGGAGT 444
QY 695 ProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsn 714
Db 443 CCTAAGTCCGTGTACACTGTGAATGTGAAGCTCCCTATGGGATTAATGTGATTGAAAG 384
QY 715 ProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTTrpPheVal 734
Db 383 CCTAAGAGGCTTGTGTAGTACGTTGATCAACAGCTTGAGCTATAGAGTTGCTTTGTG 324

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:06:22 ; Search time 46 Seconds
(without alignments)

2674.198 Million cell updates/sec

US-09-806-767-2

Title:

Perfect score: 4018

Sequence: 1 MEKPPFLCIIFLLFCSSSS.....NSHLMQVRSPISVTLKTN 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4018	100.0	775	21	AAV83300
2	4018	100.0	775	23	ABV30836
3	2786.5	69.4	766	21	AAV83304
4	2727.5	67.9	829	21	AAV83302
5	2727.5	67.9	829	21	AAV83303
6	2540	63.2	491	21	AAV83301
7	1742	43.4	754	21	AAV832219
8	1742	43.4	754	23	ABV91728
9	1700	42.3	775	21	AAG39671

10	1684	41.9	775	21	AAG28659	Arabidopsis thalia
11	1666.5	41.5	757	23	ABV33985	Herbicidally activ
12	1642.5	40.9	780	21	AAG43556	Arabidopsis thalia
13	1642.5	40.9	780	23	ABV33785	Herbicidally activ
14	1612	40.1	735	21	AAG43558	Arabidopsis thalia
15	1612	40.1	740	21	AAG43557	Arabidopsis thalia
16	1589	39.5	764	21	AAG29250	Arabidopsis thalia
17	1575.5	39.2	728	21	AAG29251	Arabidopsis thalia
18	1562	38.9	757	23	ABV90803	Herbicidally activ
19	1526.5	38.0	606	21	AAG32220	Arabidopsis thalia
20	1409.5	35.1	568	21	AAG32221	Arabidopsis thalia
21	1369	34.1	763	23	ABV38993	Herbicidally activ
22	1348.5	33.6	581	21	AAG39672	Arabidopsis thalia
23	1337.5	33.3	581	21	AAG28660	Arabidopsis thalia
24	1309	32.6	791	23	ABV37113	Herbicidally activ
25	1271.5	31.6	756	23	ABV32882	Herbicidally activ
26	1248.5	31.1	537	21	AAG39673	Arabidopsis thalia
27	1241	30.9	736	23	ABV39777	Herbicidally activ
28	1237.5	30.8	537	21	AAG28661	Arabidopsis thalia
29	1219.5	30.4	780	23	ABV91074	Herbicidally activ
30	1215.5	30.3	775	23	ABV32884	Herbicidally activ
31	1214	30.2	769	21	AAG28896	Arabidopsis thalia
32	1201.5	29.9	748	21	AAG28897	Arabidopsis thalia
33	1201.5	29.9	748	23	ABV91075	Herbicidally activ
34	1200.5	29.9	730	21	AAG28898	Arabidopsis thalia
35	1196	29.8	803	23	ABV32885	Herbicidally activ
36	1190.5	29.6	747	23	ABV32883	Herbicidally activ
37	1187.5	29.6	542	21	AAG29252	Arabidopsis thalia
38	1187.5	29.6	706	23	ABV2784	Herbicidally activ
39	1180.5	29.4	736	23	ABV32568	Herbicidally activ
40	1176	29.3	739	23	ABV32567	Herbicidally activ
41	1173	29.2	765	23	ABV32881	Herbicidally activ
42	1172.5	29.2	731	15	AAG60822	Melon protease fro
43	1145.5	28.5	746	23	ABV33111	Herbicidally activ
44	1117	27.8	772	23	ABV33038	Herbicidally activ
45	1101.5	27.4	755	23	ABV33440	Herbicidally activ

ALIGNMENTS

RESULT 1

AAV83300

ID AAV83300 standard; Protein; 775 AA.

XX

AC AAV83300;

XX

DT 16-AUG-2000 (first entry)

XX

DE SDD1 subtilisin like serine protease.

XX

KW SDD1; serine protease; subtilisin; transgenic plants; dry weight;

KW stomata; sugar; water; protein; CO 2; H 2O; CO2; H2O;

KW crop protection; feed; foodstuffs.

XX

OS Arabidopsis thaliana.

XX

PN WO200022144-A2.

XX

PD 20-APR-2000.

XX

PF 12-OCT-1999; 99WO-EP07633.

XX

PR 12-OCT-1998; 98EP-0119244.

XX

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PA (PLAN-) PLANITEC BIOTECHNOLOGIE GMBH.

XX

PI Berger D, Altmann T;

XX

DR WPI: 2000-317995/27.

DR N-PSDB; AAV33808.

XX

PT Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
PT lower water consumption and enhanced diseased resistance
XX
XX Claim 1; Page 63-65; 101pp; English.
XX
XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO₂ uptake into and H₂O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives.
XX
XX SQ Sequence 775 AA;
Query Match 100.0%; Score 4018; DB 21; Length 775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKPPFLCIIIFLLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSEFLQEAVLGV 60
DB 1 MEKPPFLCIIIFLLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSEFLQEAVLGV 60
QY 61 EEEBEPSSRLLYSGAIEGFAAQLTSEAEILRYSPVAVRDPHVLDVQVQTTYSYKFL 120
DB 61 EEEBEPSSRLLYSGAIEGFAAQLTSEAEILRYSPVAVRDPHVLDVQVQTTYSYKFL 120
QY 121 GLDGFNGSVWKSRSFGQGTIIIGVLDTGVMPEPSFDDTGMPSPRKYKWCIGQEGSFSS 180
DB 121 GLDGFNGSVWKSRSFGQGTIIIGVLDTGVMPEPSFDDTGMPSPRKYKWCIGQEGSFSS 180
QY 181 SSCNKLIGARFFIRGHRVANSPESPNPREYISARDSTGHTHTASTVGGSSVMANV 240
DB 181 SSCNKLIGARFFIRGHRVANSPESPNPREYISARDSTGHTHTASTVGGSSVMANV 240
QY 241 LGGAGVARGMAPGAHIAIVYKVCWNGCYSSDILAAIDVAIQDVLSLGLGFFIPLY 300
DB 241 LGGAGVARGMAPGAHIAIVYKVCWNGCYSSDILAAIDVAIQDVLSLGLGFFIPLY 300
QY 301 DTTAIGTFRAMERGISVTCAGNNGPIESSVANTAPWVSTTGAGTLDRRPPAVVRLANG 360
DB 301 DTTAIGTFRAMERGISVTCAGNNGPIESSVANTAPWVSTTGAGTLDRRPPAVVRLANG 360
QY 361 KLLYGSLYPGKIGKNAGREVEVIVYTGDKGSEFLCRLSPREIRGKMWICDRGVNGR 420
DB 361 KLLYGSLYPGKIGKNAGREVEVIVYTGDKGSEFLCRLSPREIRGKMWICDRGVNGR 420
QY 421 SEKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNATVKPKAR 480
DB 421 SEKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNATVKPKAR 480
QY 481 IIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPQNLGPTGLPYDS 540
DB 481 IIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPQNLGPTGLPYDS 540
QY 541 RRNVFTVMSGTSMSCPHVSGITALLRSAYPNMSPAAIKSALMTTADLYDRQKAIKQNGK 600
DB 541 RRNVFTVMSGTSMSCPHVSGITALLRSAYPNMSPAAIKSALMTTADLYDRQKAIKQNGK 600
QY 601 PAGVFAIGAGHVNPKAINPGLVYNIQPDYIITYLCTLGTFRSDILAITHKVNSCNGILR 660
DB 601 PAGVFAIGAGHVNPKAINPGLVYNIQPDYIITYLCTLGTFRSDILAITHKVNSCNGILR 660
QY 661 KNPGLSLNPSIAVIFKRGKTTMIRRTVNVGSPNSIYSVNVKAPEGIKVINPKRLVF 720
DB 661 KNPGLSLNPSIAVIFKRGKTTMIRRTVNVGSPNSIYSVNVKAPEGIKVINPKRLVF 720
QY 721 KHVDTQLSYRVWVFLKKKVRGGKVASFAQGQLTWNVSHNLMQVRSPISVTLKTN 775

DB 721 KHVDTQLSYRVWVFLKKKVRGGKVASFAQGQLTWNVSHNLMQVRSPISVTLKTN 775
RESULT 2
ABS90836
ID ABS90836 standard; Protein; 775 AA.
XX
XX AC ABS90836;
XX
XX DT 31-MAY-2002 (first entry)
XX
XX DE Herbicidally active polypeptide SEQ ID NO 47.
XX
XX DE Herbicidal; plant; agriculture; herbicide.
XX
XX KW Arabidopsis thaliana.
XX
XX OS Arabidopsis thaliana.
XX
XX PN WO200210210-A2.
XX
XX PD 07-FEB-2002.
XX
XX PF 28-AUG-2001; 2001WO-EP09892.
XX
XX PR 28-AUG-2001; 2001WO-EP09892.
XX
XX PA (FARB) BAYER AG.
XX
XX PI Tietjen K, Weidler M;
XX
XX DR WPI; 2002-269010/31.
XX
XX PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
XX PS Claim 5; SEQ ID NO 47; 261pp + Sequence Listing; English.
XX
XX CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
XX SQ Sequence 775 AA;
Query Match 100.0%; Score 4018; DB 23; Length 775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKPPFLCIIIFLLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSEFLQEAVLGV 60
DB 1 MEKPPFLCIIIFLLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSEFLQEAVLGV 60
QY 61 EEEBEPSSRLLYSGAIEGFAAQLTSEAEILRYSPVAVRDPHVLDVQVQTTYSYKFL 120
DB 61 EEEBEPSSRLLYSGAIEGFAAQLTSEAEILRYSPVAVRDPHVLDVQVQTTYSYKFL 120
QY 121 GLDGFNGSVWKSRSFGQGTIIIGVLDTGVMPEPSFDDTGMPSPRKYKWCIGQEGSFSS 180
DB 121 GLDGFNGSVWKSRSFGQGTIIIGVLDTGVMPEPSFDDTGMPSPRKYKWCIGQEGSFSS 180
QY 181 SSCNKLIGARFFIRGHRVANSPESPNPREYISARDSTGHTHTASTVGGSSVMANV 240
DB 181 SSCNKLIGARFFIRGHRVANSPESPNPREYISARDSTGHTHTASTVGGSSVMANV 240
QY 241 LGGAGVARGMAPGAHIAIVYKVCWNGCYSSDILAAIDVAIQDVLSLGLGFFIPLY 300
DB 241 LGGAGVARGMAPGAHIAIVYKVCWNGCYSSDILAAIDVAIQDVLSLGLGFFIPLY 300


```

QY 301 DDTIAIGTFRAMERGI:SVICAAAGNGPIESSVANTAPWVSTTICAGTLDRRFFPAVVRANG 360
DB 301 DDTIAIGTFRAMERGI:SVICAAAGNGPIESSVANTAPWVSTTICAGTLDRRFFPAVVRANG 360
QY 361 KLLYGESLYPGKIGKNAGREVEVYVYTGDKGSEFCRLRGLPREIRGQWVICDRGVNGR 420
DB 361 KLLYGESLYPGKIGKNAGREVEVYVYTGDKGSEFCRLRGLPREIRGQWVICDRGVNGR 420
QY 421 SEKGBAVKEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKAYVNATVKPKAR 480
DB 421 SEKGBAVKEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKAYVNATVKPKAR 480
QY 481 IIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPONLGTGLPYDS 540
DB 481 IIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPONLGTGLPYDS 540
QY 541 RRVNFTVMSGTSMSCHVSGITAIRSAYPNPSPAIAKALMTTADLYDROGKAIDGNK 600
DB 541 RRVNFTVMSGTSMSCHVSGITAIRSAYPNPSPAIAKALMTTADLYDROGKAIDGNK 600
QY 601 PAGVFAIGAGHVPKAINPGLVYN:OPVDY:TYLCTLGFTSRSDILAITHKNVSCNGILR 660
DB 601 PAGVFAIGAGHVPKAINPGLVYN:OPVDY:TYLCTLGFTSRSDILAITHKNVSCNGILR 660
QY 661 KNPFGSLNYPESIAVIFKRGKTTMIRRTVNVGSPNSIYSVNKAPGEG:KVIWNPRLVF 720
DB 661 KNPFGSLNYPESIAVIFKRGKTTMIRRTVNVGSPNSIYSVNKAPGEG:KVIWNPRLVF 720
QY 721 KHVDTQLSVRVWFVLLKKRGGKVASFAOGLTWNVSHNLMQVRSPISVTILKTN 775
DB 721 KHVDTQLSVRVWFVLLKKRGGKVASFAOGLTWNVSHNLMQVRSPISVTILKTN 775

```

RESULT 3

```

AAV83304
ID AAV83304 standard; Protein; 766 AA.
XX AAV83304;
AC AAV83304;
DT 16-AUG-2000 (first entry)
XX Subtilase (St_28Sc) of Solanum tuberosum.
XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomata; sugar; water; protein; CO 2; H 2O; CO2; H2O;
KW crop protection; feed; foodstuffs; ss.
XX Solanum tuberosum.
OS Solanum tuberosum.

```

```

XX Key Location/Qualifiers
FH Misc-difference 729
FT /note= "Unidentified amino acid"
XX
XX W0200022144-A2.
XX
XX 20-APR-2000.
XX
XX 12-OCT-1999; 99WO-EP07633.
XX
XX 12-OCT-1998; 98EP-0119244.
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
XX Berger D, Altmann T;
XX
XX WPI; 2000-317995/27.
DR N-PSDB; AA293814.
XX
XX Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
PT lower water consumption and enhanced diseased resistance

```

```

XX
PS
XX
CC Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristic. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO 2 uptake into and H 2O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This subfamily of
CC Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
CC (See AY83300).
XX
SQ Sequence 766 AA;

```

```

Query Match 69.4%; Score 2786.5; DB 21; Length 766;
Best Local Similarity 68.5%; Pred. No. 3.5e-227;
Matches 524; Conservative 107; Mismatches 125; Indels 9; Gaps 5;

```

```

QY 7 FLCIFLLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSPFOEAVLGVVEEBEE 66
DB 7 FLCIFLLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSPFOEAVLGVVEEBEE 66
QY 67 PSSRLLYSGSATEGFAAQLTSEAEILRYSPVAVRPHVLQVQTYYSKYLGLDGF 126
DB 67 PSSRLLYSGSATEGFAAQLTSEAEILRYSPVAVRPHVLQVQTYYSKYLGLDGF 126
QY 60 SSSRLLYSYHSAMEGFAARLTDEVELLRSDNVLISRAERRLIEIQTYSYKFLGSP-T 118
DB 60 SSSRLLYSYHSAMEGFAARLTDEVELLRSDNVLISRAERRLIEIQTYSYKFLGSP-T 118
QY 127 NSGVMSXSRFGQGTIIIGVLDTGVMPEPSPDDTGMPISIPRKWKIGICEGSEFSSSSNRK 186
DB 127 NSGVMSXSRFGQGTIIIGVLDTGVMPEPSPDDTGMPISIPRKWKIGICEGSEFSSSSNRK 186
QY 119 REGAWLXSGFGRGAIIGVLDTGVMPEPSPDDTGMPISIPRKWKIGICEGSEFSSSSNRK 178
DB 119 REGAWLXSGFGRGAIIGVLDTGVMPEPSPDDTGMPISIPRKWKIGICEGSEFSSSSNRK 178
QY 187 LIGARFFIRGHRVANSPEESPNNPREYISARDSTGHTGHTTASTVGGSSVSMANVLNGAG 246
DB 187 LIGARFFIRGHRVANSPEESPNNPREYISARDSTGHTGHTTASTVGGSSVSMANVLNGAG 246
QY 179 LIGARFFIRGHRVANSPEESPNNPREYISARDSTGHTGHTTASTVGGSSVSMANVLNGAG 237
DB 179 LIGARFFIRGHRVANSPEESPNNPREYISARDSTGHTGHTTASTVGGSSVSMANVLNGAG 237
QY 247 VARGNAPGAHTAVYKVCWFNGCYSSDILAAIDVAIQDKVDLSLGGPFIPLYDDTIAI 306
DB 247 VARGNAPGAHTAVYKVCWFNGCYSSDILAAIDVAIQDKVDLSLGGPFIPLYDDTIAI 306
QY 238 EARGWAPGAHTAVYKVCWFNGCYSSDILAAIDVAIQDKVDLSLGGPFIPLYDDTIAI 297
DB 238 EARGWAPGAHTAVYKVCWFNGCYSSDILAAIDVAIQDKVDLSLGGPFIPLYDDTIAI 297
QY 307 GTFRAMERGISVICAAGNGPIESSVANTAPWVSTTICAGTLDRRFFPAVVRANGKLYGE 366
DB 307 GTFRAMERGISVICAAGNGPIESSVANTAPWVSTTICAGTLDRRFFPAVVRANGKLYGE 366
QY 298 GSFRAMEHGISVICAAGNGPIESSVANGAPWATIIICASTLDRFPASVQNGKFLYGE 357
DB 298 GSFRAMEHGISVICAAGNGPIESSVANGAPWATIIICASTLDRFPASVQNGKFLYGE 357
QY 367 SLYPGKIGKNAGREVEVYVYTGDKGSEFCRLRGLPREIRGQWVICDRGVNGRSEKGEA 426
DB 367 SLYPGKIGKNAGREVEVYVYTGDKGSEFCRLRGLPREIRGQWVICDRGVNGRSEKGEA 426
QY 358 SLYPGKVPSSQKLE:VYVKDKDKGSEFCRLRGLSKAQVRGVNVCVRGVNRAEKGQV 417
DB 358 SLYPGKVPSSQKLE:VYVKDKDKGSEFCRLRGLSKAQVRGVNVCVRGVNRAEKGQV 417
QY 427 VKEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKAYVNATVKPKARIIFGGT 486
DB 427 VKEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKAYVNATVKPKARIIFGGT 486
QY 418 VKEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKAYVNATVKPKARIIFGGT 477
DB 418 VKEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKAYVNATVKPKARIIFGGT 477
QY 487 VIGSRAPVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPONLGTGLPYDSRRVNYFT 546
DB 487 VIGSRAPVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPONLGTGLPYDSRRVNYFT 546
QY 478 VIGSRAPVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPONLGTGLPYDSRRVNYFT 537
DB 478 VIGSRAPVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPONLGTGLPYDSRRVNYFT 537
QY 547 VMSGTSMCHVSGITAIRSAYPNPSPAIAKALMTTADLYDROGKAIDGNKPAVFA 606
DB 547 VMSGTSMCHVSGITAIRSAYPNPSPAIAKALMTTADLYDROGKAIDGNKPAVFA 606
QY 538 VMSGTSMCHVSGITAIRSAYPNPSPAIAKALMTTADLYDROGKAIDGNKPAVFA 597
DB 538 VMSGTSMCHVSGITAIRSAYPNPSPAIAKALMTTADLYDROGKAIDGNKPAVFA 597
QY 607 IGAGHVPKAINPGLVYN:OPVDY:TYLCTLGFTSRSDILAITHKNVSCNGILKPNPFS 666
DB 607 IGAGHVPKAINPGLVYN:OPVDY:TYLCTLGFTSRSDILAITHKNVSCNGILKPNPFS 666
QY 598 AGAGHVPKAINPGLVYN:OPVDY:TYLCTLGFTSRSDILAITHKNVSCNGILKPNPFS 657
DB 598 AGAGHVPKAINPGLVYN:OPVDY:TYLCTLGFTSRSDILAITHKNVSCNGILKPNPFS 657
QY 667 LNPYSIAVIFKRGKTTMIRRTVNVGSPNSIYSVNKAPGEG:KVIWNPRLVF 726
DB 667 LNPYSIAVIFKRGKTTMIRRTVNVGSPNSIYSVNKAPGEG:KVIWNPRLVF 726
QY 658 LNPYSIAVIFKRGKTTMIRRTVNVGSPNSIYSVNKAPGEG:KVIWNPRLVF 717
DB 658 LNPYSIAVIFKRGKTTMIRRTVNVGSPNSIYSVNKAPGEG:KVIWNPRLVF 717
QY 727 LSVYVWFVLLKKRGGKVASFAOGLTWNVSHNLMQVRSPISVT 771
DB 727 LSVYVWFVLLKKRGGKVASFAOGLTWNVSHNLMQVRSPISVT 771
QY 718 LSVYVWFVLLKKRGGKVASFAOGLTWNVSHNLMQVRSPISVT 761
DB 718 LSVYVWFVLLKKRGGKVASFAOGLTWNVSHNLMQVRSPISVT 761

```



```

PR 12-OCT-1998; 98EP-0119244.
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
XX Berger D, Altmann T;
XX
XX WPI; 2000-317995/27.
XX DR N-PSDB; AAZ93813.
XX
XX Novel recombinant DNA molecules encoding subtilisin-like serine
XX protease, useful for producing transgenic plants with altered stomata,
XX lower water consumption and enhanced diseased resistance
XX
XX Claim 1; Page 82-84; 101pp; English.
XX
XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
XX used to produce transgenic plants with altered stomata
XX characteristics. These plants exhibit improved freshness,
XX increased dry weight, reduced leaf temperatures, reduced water loss
XX and lower water consumption and for enhancing the sugar and/or
XX protein content of plant leaves, modulating CO2 uptake into and H2O
XX release from leaves, for sustained photosynthesis under high
XX intensity conditions or for the improvement of disease resistance
XX of plants. The transgenic plants and cells of such plants are useful
XX in the preparation of feed, food or additives. This subtitase of
XX Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
XX (See AAY83300).
XX
XX Sequence 829 AA;
XX
XX Query Match 67.9%; Score 2727.5; DB 21; Length 829;
XX Best Local Similarity 65.6%; Pred. No. 4e-222;
XX Matches 521; Conservative 108; Mismatches 132; Indels 33; Gaps 5;
XX
XX 6 FFLCIIFLFCSSSSBILOKQYIVQLHPNSETAKT-PASKEDWHLFLQE----- 55
XX 34 YFLC--FLLCFTPLLAQNLQYIVQLHPHASTRTPFSSKQWHLSPLENTFNPLNF 91
XX
XX 56 -----AVLGEDEEPEPSRLLYSGSAIEGFAAQLTESEAIIRY 96
XX 92 KYIQWNSIFILFCYVSYPATSSISGENSSRLLYSYHSAFEGFAALLSENELKALK 151
XX
XX 97 SPBVAVRPHVLQVCTTYSYFLGLDGFNGSVKSKRFGGTIIIGVLDTCWGPSPSF 156
XX 152 SNNVLSIYPERKLEVTYTYKPLGSP-TKSGTMDXSGFGGAIIGVLDTCIWGPSPSF 210
XX
XX 157 DDTGMPISIPKWKIGICQEGESFSSSCNKLICARPPFIRGHRVANSPEPSNMPREISA 216
XX 211 VDHGMSPIPKWKXGXCQEGNFSSSCNKLICARPPFQIGHMMASKTSIDFMDYSP 270
XX
XX 217 RDSGTGHTTASTVGGSSVSMANVLGNAGVARGMAPGAHIAVYKCVWFGYSSDILAA 276
XX 271 RDSQGHGHTTASTAGCAPVPMAVLGNAGARGMAPGAHIAIYKCVSSGCGYSSDILAA 330
XX
XX 277 IDVAIQKVDVLSLGGPPIPLDYDTIAGTFRAMERGISVICAGNNGPIESSVANTA 336
XX 331 MDVAIRDGVLDLSLGGPVPPLDYDTIAGTFRAMERGISVICAGNNGPILSSVANE 390
XX 337 PWYSTTGAGTLDRRFPVAVRLKLLYGESIYPGKIKNAGREVEIVVTGDDGSGSFC 396
XX 391 PWATIGASTLDRKFAITQLGNKGYGESIYPGQVHNSQKLEIVTLNDGNGSGSFC 450
XX
XX 397 LRGLSPREIRGHWICDRGVNRSKGAHVKAAGVAMILANTINOEDSIDVHLLPA 456
XX 451 LRGLSPRAKHGKIIVCDRGVNGRAEKQGVKXESGGVAMILANTAVNNEEDSDVHLLPA 510
XX
XX 457 TLIGYTESVLLKAYNATVKPARIIPGTGTIGRSPAPVAFSARGSPANPSILKPD 516
XX 511 TLIGFDESILQSYMNSRKPTARIIFGCTVIGKSAPAVAFSGRSPFTDPSILKPDV 570
XX
XX 517 IAPGVNIIAAMPONLQGLPYDSRVNFTVMSGFSMSCPHVSIGTALIRSAFYNWSPAA 576
XX

```

```

Db 571 IAPGVNIIAAMPONLQGLPGLAEDSRVRNFTVLSGTSMACHVSGIAALLHSHPKWSPAA 630
Qy 577 IKSALMTTADLYORQKAIKIDGNKPAQFAIGAGHVNPQKAINPGLVYNIQPYDYITL 636
Db 631 IKSALMTTADTTHQKQKPIINDGTRAGLFAIGAGHVNPQKSGDDPGLIYDINANDYITHLC 690
Qy 637 TLGFTSRDILAIATHKNVSCNGIILKQKPGFSLNVPSPSTAVIPKRGKTTMTRVTNVGSPN 696
Db 691 TIGYKNSIILSIHKVNSCHDLQKRGFSLNVPSPSTAVIPKRGKTRMTRVTNVGSPN 750
Qy 697 SIYSVNVKAPGKIVIVNPKRLVFKVDOTLSYRVNFKKNGKGVKVASPAQGLTWN 756
Db 751 SIYSVEIVAPGKVRKVRPRRLVFKVNSQSLSYRVNFI-SRKRIQTQRSPFAGQLMWIN 809
Qy 757 SHNLMQVRSPISV 770
Db 810 SRDKYQKVRSPISV 823
XX
XX RESULT 6
XX ID AAY83301 standard; Protein; 491 AA.
XX AC AAY83301;
XX AC AAY83301;
XX DT 16-AUG-2000 (first entry)
XX DE Prematurely terminated SDD1 subtilisin like serine protease.
XX
XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
XX stomata; sugar; water; protein; CO2; H2O; CO2; H2O;
XX crop protection; feed; foodstuffs.
XX
XX Arabidopsis thaliana.
XX
XX WO2000022144-A2.
XX
XX 20-APR-2000.
XX
XX 12-OCT-1999; 99WO-EP07633.
XX
XX 12-OCT-1998; 98EP-0119244.
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
XX Berger D, Altmann T;
XX
XX WPI; 2000-317995/27.
XX N-PSDB; AAZ93809.
XX
XX Novel recombinant DNA molecules encoding subtilisin-like serine
XX protease, useful for producing transgenic plants with altered stomata,
XX lower water consumption and enhanced diseased resistance
XX
XX Claim 1; Page 68-70; 101pp; English.
XX
XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
XX used to produce transgenic plants with altered stomata
XX characteristics. These plants exhibit improved freshness,
XX increased dry weight, reduced leaf temperatures, reduced water loss
XX and lower water consumption and for enhancing the sugar and/or
XX protein content of plant leaves, modulating CO2 uptake into and H2O
XX release from leaves, for sustained photosynthesis under high
XX intensity conditions or for the improvement of disease resistance
XX of plants. The transgenic plants and cells of such plants are useful
XX in the preparation of feed, food or additives. This sequence is
XX identical to the one given in GENESQ record AAY83300 with the
XX exception that a single point mutation in the coding sequence
XX introduces a TGA codon and results in premature termination
XX of the SDD1 polypeptide.
XX
XX Sequence 491 AA;
XX

```

```
Query Match          63.2%; Score 2540; DB 21; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.4e-206;
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPKPFLCIIIFLLCCSSSEILQQTIVIVQLHPNSETAKTFASKFDWHLFLQEAVLGV 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MEPKPFLCIIIFLLCCSSSEILQQTIVIVQLHPNSETAKTFASKFDWHLFLQEAVLGV 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 EEEEEPSSRLIYSGSAIEGFAAOLTESEARILYSEVAVRDPDHVLQVTTYSYKFL 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 EEEEEPSSRLIYSGSAIEGFAAOLTESEARILYSEVAVRDPDHVLQVTTYSYKFL 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 121 GLDGFNSGVNSKSFQGTIIIGVLDTGWPSPSFDGTGMPSPRKKWGIQQGESEFSS 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 GLDGFNSGVNSKSFQGTIIIGVLDTGWPSPSFDGTGMPSPRKKWGIQQGESEFSS 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 181 SSCNRKLI GARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVMANV 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 SSCNRKLI GARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVMANV 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 241 LONGAGVAGMAPAGAHIAVYKVCWNGCYSSDILAAIDVAIQDKVDVLSLSLGGFPPIPLY 300
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 LONGAGVAGMAPAGAHIAVYKVCWNGCYSSDILAAIDVAIQDKVDVLSLSLGGFPPIPLY 300
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 301 DDTIAGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLANG 360
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 DDTIAGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLANG 360
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 361 KLLYGESLYPGKIGKNAGREVEVIYVTGDKGSEFCLRGSLPREIRKMWICDRGVNGR 420
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 KLLYGESLYPGKIGKNAGREVEVIYVTGDKGSEFCLRGSLPREIRKMWICDRGVNGR 420
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 421 SKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNATVKKPKAR 480
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 421 SKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNATVKKPKAR 480
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 481 IIFGGTVIGRS 491
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 481 IIFGGTVIGRS 491
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
AAG32219 ID AAG32219 standard; Protein; 754 AA.
XX AC AAG32219;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 38826.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125786.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 05-MAY-1999; 99US-0132484.
XX PR 06-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139750.
XX PR 18-JUN-1999; 99US-0139763.
XX PR 21-JUN-1999; 99US-0139817.
XX PR 22-JUN-1999; 99US-0139899.
XX PR 23-JUN-1999; 99US-0140353.
XX PR 23-JUN-1999; 99US-0140354.
XX PR 24-JUN-1999; 99US-0140695.
XX PR 28-JUN-1999; 99US-0140823.
XX PR 29-JUN-1999; 99US-0140991.
XX PR 30-JUN-1999; 99US-0141287.
XX PR 01-JUL-1999; 99US-0141842.
XX PR 02-JUL-1999; 99US-0142154.
XX PR 02-JUL-1999; 99US-0142055.
XX PR 06-JUL-1999; 99US-0142390.
XX PR 08-JUL-1999; 99US-0142803.
XX PR 09-JUL-1999; 99US-0142920.
XX PR 12-JUL-1999; 99US-0142977.
XX PR 13-JUL-1999; 99US-0143542.
XX PR 14-JUL-1999; 99US-0143624.
XX PR 15-JUL-1999; 99US-0144005.
XX PR 16-JUL-1999; 99US-0144085.
XX PR 16-JUL-1999; 99US-0144086.
XX PR 19-JUL-1999; 99US-0144325.
XX PR 19-JUL-1999; 99US-0144331.
XX PR 19-JUL-1999; 99US-0144332.
XX PR 19-JUL-1999; 99US-0144333.
```

```
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 43.4%; Score 1742; DB 21; Length 754;
Best Local Similarity 48.4%; Pred. No. 1.5e-138;
Matches 370; Conservative 118; Mismatches 237; Indels 40; Gaps 16;

QY 11 IFLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSPLOEAVLGVVEEPEPSSR 70
Db 18 LFLLLHTTA-----KKTYILIRVN-HSKPESFLTHDWTYSQLNS-----ESS 59

QY 71 LLYSGSAIEGPAQLTESPAE-ILRYSPEVAVRPHVLQVQTYVSKYKFLGLDG-FCNS 128
Db 60 LLXTYTTSPHGFAYLSDTEADSLSSNSILDIPELPLTYLTHTRTPEFLGNSPFGVH 119

QY 129 GVMKSRFGOGTIIIGVLDTCGWPEPSFDDTGMPSPRKWKGCQEGESFSSSCNRKLI 188
Db 120 DLGSSS---NGVIIGVLDTCGWPEPSFDDTGMPSPRKWKGECSGSDPDKLCKNKLI 176

QY 189 GARFFIRGHRVANSPEPSNPNPREYISARDSTGHTGHTASTVGSSSVSMANVLGNGAGVA 248
Db 177 GARFSKGFQWASGGFSSK--RESVSPRDVDGHTHTSTTAAGSAVRNASFLGYAAGTA 234

QY 249 RGMAPGAHIAVYKWFNGCYSSDILAAIDVAIOKVDVLSLSLGGFFPIPLYDDTTAIGT 308
Db 235 RGMATRARVATYKVCNSTGCGFSDILAAAMDRAILDGVDVLSLSLGGCSAPYRDITAI 294

QY 309 FRAMERGISVICAAGNNGPIESSVANTAPWVTIGAGTILDRRFPVAVRVLANKLLYGESL 368
Db 295 FSAMERGVFVSCSAGNSGPTRASVANVAPWMTVGAGTILDRDPFAPANLNGKRLTGVS 354

QY 369 YPGKGIKNAGREVEVIYVTGDKGSEFCRLGRSPREIRGKVMICDRGVNGRSEKGEAVK 428
Db 355 YSGVGM--GTRKPLELVYKNGKNSSSNLCPLFESLDSISVRGKI VVCDRGVNARVEKAVR 412

QY 429 EAGGVAMILANTEINOBEEDSIDVHLLPATLIGYTESVLLKAYVNAVTKPKARIIFGOTVI 488
Db 413 DAGOLGMIMANTAASGEELVADSHLLPAIAVGKKTGDLIREYVYKSDSKPTALLVFKGT 472

QY 489 GRSRAPEVAOFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGLPYDSRVNPTVM 548
Db 473 DVKPSVVAAFSSRGENTVTPKILKPDVIGVGNILAGWSDAICPTGLDKDSRRTQNM 532

QY 549 SGTSMSCPHVSGITALIRSAIPNKSAPAAIKSALMTADLYDROGKAIKDG--NKPAGVFA 606
Db 533 SGTSMSCPHISGLAGLLKAAHPWSPSAIKSALMTAYVLDNTNAPLHDAADSLNSNPIYA 592

QY 607 IGAGHVNPQKAINPGLVYNIQPVDIYTYLCTLGTFRSDILAI THK-NVSCNGILRNKPGF 665
Db 593 HSGSHVDPOKALSPGLYDIDISTEYIRFLCSLYTYVDHIVAIKVRPSVNCSSKXF-SDPG- 650
```


PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 25-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0133456.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134421.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148694.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155119.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157885.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.


```
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.3%; Score 1700; DB 21; Length 775;
Best Local Similarity 45.0%; Pred. No. 5.8e-135;
Matches 356; Conservative 134; Mismatches 257; Indels 44; Gaps 15;

QY 6 FFLCTIIFLLFCSSSEILQKQ---TYIVQLHPNSETAKTFASKFDWHLSPLOEAVLGVBE 62
DB 3 FFFVFFFLITLSSPSSASSNSLTYIVHVDHEAKPS-IPTHFHWYSSLASLT----- 56

QY 63 EEEPSRLLYSYGAIEGFAAQTESEAILRYSPEVAVRPHVLQVOTTSYKFLGL 122
DB 57 --SSPPS-IITYDTPVHGFSARLTQSDASQLLDHPHVISIPEQVRHLHTTSPFGL 113

QY 123 DFGNGSWKSRGQGTIIGVLTGWPSPSPDDTGMPSPKXKGIQCEGESFSSSS 182
DB 114 RSTKAGLEESDFGSLVIGVIDTGVNPERPDDRGLGFVPKWKQGCIAQDPESA 173

QY 183 CNRKLIGARFIRHRVANSPEESPNNPRVISAARDSTGHTGTASTVGGSSVSMANVLG 242
DB 174 CNRKLVGARFCGGYEATNG---KXNETTFRSPDSGDGHTTASISAGRYVFPASTLG 230

QY 243 NGAVGARMAPGAHIAVKVWENGSCYSDILAAIDVAIOKVDVJSLSLGGFPPIPLYDD 302
DB 231 YAHGVAAGMAPKARLAAYKVCWNSGCYDSDILAAFDTAADGVDDVISLSVGVVVPYLD 290

QY 303 TIAGTFRAMERGISVICAAGNGPIESSVANTAPWVSTIGAGTLDRRPPAVVRLANGKL 362
DB 291 AIAIGAFGAIDRGI FVSASAGNGGFGALTVTNAPWMTTVGAGTIDRDFPANKLGNKGM 350

QY 363 LYGESLYPGKIKNAGEVEVIY---VTGDD-KGSEFCLEGLSPREIRKGVICDRGVN 418
DB 351 ISGVSVYGGPL-DPGRWYPLVYGGSLGGYSSSLCLEGLDPLNVRKIVLCRGIN 409

QY 419 GRSEKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNAVK-- 476
DB 410 SRATKGEIVRKNGLGLMIANGVDGGLVADCHVLPAATSVGASGGDEIRYISESKSR 469

QY 477 ----PKAIIFGGTIGRSRAPEVAQFSARGPSIANPSILKPDMAFGVNIIAAPQNLG 532
DB 470 SSKHPTATVFKGRLGIRPAPVAVSFARGPNPETPEILKPDVAPGLNLAAPDRIG 529

QY 533 PTGLPYDSRRVNTVMSTGSCSVHSGITALLISAYPNWSPAIAKSMATTDLYDRQG 592
DB 530 PSQVTSDDRRTEFNLGTSWACPHVSGLAALKAAHPDWSPAIRSAITTYATVDNSG 589

QY 593 KAIKD---GNKPAGVFAIGAGHVNPKAINPLGVNIOPVDYITVLTGFTSRDILAIT 649
DB 590 EPMWDESTGN-TSSVMDYSGSHVHTKAMDPLGVIDITSYDIYNFLCNSNYTRINVTIT 648
```

```
QY 650 HKNVSCNGILRKNPGFSLNTPSIAVIPKRGKTTMIT---RRVTNUGSPNSIYSVNVKAP 706
DB 649 RRQADCDGARRAGHVGNLNTPSFVWFQYQYCESQNSTHFIRVTNUGSDSDSVYEIKIRPP 708
QY 707 EGKIVVNPKELVFKHYDQTLISYRVWFLKKQKRGKVASFA-----QGQLTWNSHNLQ 762
DB 709 RGTTVTEPEKLSFRRVGQKLS----FVVRVKTTVKLSPGATNVETGHIWSDG---KR 761
QY 763 RVRSPISVTLK 773
DB 762 NVTSPVLVTLQ 772

RESULT 10
AAG28659
ID AAG28659 standard; Protein; 775 AA.
XX
AC AAG28659;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 33961.
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX Arabidopsis thaliana.
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
```

PR 07-JUN-1999; 99US-01337724.
PR 08-JUN-1999; 99US-01380094.
PR 09-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 11-JUN-1999; 99US-0139119.
PR 12-JUN-1999; 99US-0139452.
PR 13-JUN-1999; 99US-0139453.
PR 14-JUN-1999; 99US-0139454.
PR 15-JUN-1999; 99US-0139455.
PR 16-JUN-1999; 99US-0139456.
PR 17-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 19-JUN-1999; 99US-0139459.
PR 20-JUN-1999; 99US-0139460.
PR 21-JUN-1999; 99US-0139461.
PR 22-JUN-1999; 99US-0139462.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 25-JUN-1999; 99US-0140355.
PR 26-JUN-1999; 99US-0140356.
PR 27-JUN-1999; 99US-0140357.
PR 28-JUN-1999; 99US-0140358.
PR 29-JUN-1999; 99US-0140359.
PR 30-JUN-1999; 99US-0140360.
PR 01-JUL-1999; 99US-0140361.
PR 02-JUL-1999; 99US-0140362.
PR 03-JUL-1999; 99US-0140363.
PR 04-JUL-1999; 99US-0140364.
PR 05-JUL-1999; 99US-0140365.
PR 06-JUL-1999; 99US-0140366.
PR 07-JUL-1999; 99US-0140367.
PR 08-JUL-1999; 99US-0140368.
PR 09-JUL-1999; 99US-0140369.
PR 10-JUL-1999; 99US-0140370.
PR 11-JUL-1999; 99US-0140371.
PR 12-JUL-1999; 99US-0140372.
PR 13-JUL-1999; 99US-0140373.
PR 14-JUL-1999; 99US-0140374.
PR 15-JUL-1999; 99US-0140375.
PR 16-JUL-1999; 99US-0140376.
PR 17-JUL-1999; 99US-0140377.
PR 18-JUL-1999; 99US-0140378.
PR 19-JUL-1999; 99US-0140379.
PR 20-JUL-1999; 99US-0140380.
PR 21-JUL-1999; 99US-0140381.
PR 22-JUL-1999; 99US-0140382.
PR 23-JUL-1999; 99US-0140383.
PR 24-JUL-1999; 99US-0140384.
PR 25-JUL-1999; 99US-0140385.
PR 26-JUL-1999; 99US-0140386.
PR 27-JUL-1999; 99US-0140387.
PR 28-JUL-1999; 99US-0140388.
PR 29-JUL-1999; 99US-0140389.
PR 30-JUL-1999; 99US-0140390.
PR 01-AUG-1999; 99US-0140391.
PR 02-AUG-1999; 99US-0140392.
PR 03-AUG-1999; 99US-0140393.
PR 04-AUG-1999; 99US-0140394.
PR 05-AUG-1999; 99US-0140395.
PR 06-AUG-1999; 99US-0140396.

PR 06-AUG-1999; 99US-0140397.
PR 07-AUG-1999; 99US-0140398.
PR 08-AUG-1999; 99US-0140399.
PR 09-AUG-1999; 99US-0140400.
PR 10-AUG-1999; 99US-0140401.
PR 11-AUG-1999; 99US-0140402.
PR 12-AUG-1999; 99US-0140403.
PR 13-AUG-1999; 99US-0140404.
PR 14-AUG-1999; 99US-0140405.
PR 15-AUG-1999; 99US-0140406.
PR 16-AUG-1999; 99US-0140407.
PR 17-AUG-1999; 99US-0140408.
PR 18-AUG-1999; 99US-0140409.
PR 19-AUG-1999; 99US-0140410.
PR 20-AUG-1999; 99US-0140411.
PR 21-AUG-1999; 99US-0140412.
PR 22-AUG-1999; 99US-0140413.
PR 23-AUG-1999; 99US-0140414.
PR 24-AUG-1999; 99US-0140415.
PR 25-AUG-1999; 99US-0140416.
PR 26-AUG-1999; 99US-0140417.
PR 27-AUG-1999; 99US-0140418.
PR 28-AUG-1999; 99US-0140419.
PR 29-AUG-1999; 99US-0140420.
PR 30-AUG-1999; 99US-0140421.
PR 31-AUG-1999; 99US-0140422.
PR 01-SEP-1999; 99US-0140423.
PR 02-SEP-1999; 99US-0140424.
PR 03-SEP-1999; 99US-0140425.
PR 04-SEP-1999; 99US-0140426.
PR 05-SEP-1999; 99US-0140427.
PR 06-SEP-1999; 99US-0140428.
PR 07-SEP-1999; 99US-0140429.
PR 08-SEP-1999; 99US-0140430.
PR 09-SEP-1999; 99US-0140431.
PR 10-SEP-1999; 99US-0140432.
PR 11-SEP-1999; 99US-0140433.
PR 12-SEP-1999; 99US-0140434.
PR 13-SEP-1999; 99US-0140435.
PR 14-SEP-1999; 99US-0140436.
PR 15-SEP-1999; 99US-0140437.
PR 16-SEP-1999; 99US-0140438.
PR 17-SEP-1999; 99US-0140439.
PR 18-SEP-1999; 99US-0140440.
PR 19-SEP-1999; 99US-0140441.
PR 20-SEP-1999; 99US-0140442.
PR 21-SEP-1999; 99US-0140443.
PR 22-SEP-1999; 99US-0140444.
PR 23-SEP-1999; 99US-0140445.
PR 24-SEP-1999; 99US-0140446.
PR 25-SEP-1999; 99US-0140447.
PR 26-SEP-1999; 99US-0140448.
PR 27-SEP-1999; 99US-0140449.
PR 28-SEP-1999; 99US-0140450.
PR 29-SEP-1999; 99US-0140451.
PR 30-SEP-1999; 99US-0140452.
PR 01-OCT-1999; 99US-0140453.
PR 02-OCT-1999; 99US-0140454.
PR 03-OCT-1999; 99US-0140455.
PR 04-OCT-1999; 99US-0140456.
PR 05-OCT-1999; 99US-0140457.
PR 06-OCT-1999; 99US-0140458.
PR 07-OCT-1999; 99US-0140459.
PR 08-OCT-1999; 99US-0140460.
PR 09-OCT-1999; 99US-0140461.
PR 10-OCT-1999; 99US-0140462.
PR 11-OCT-1999; 99US-0140463.
PR 12-OCT-1999; 99US-0140464.
PR 13-OCT-1999; 99US-0140465.
PR 14-OCT-1999; 99US-0140466.
PR 15-OCT-1999; 99US-0140467.
PR 16-OCT-1999; 99US-0140468.
PR 17-OCT-1999; 99US-0140469.
PR 18-OCT-1999; 99US-0140470.
PR 19-OCT-1999; 99US-0140471.
PR 20-OCT-1999; 99US-0140472.
PR 21-OCT-1999; 99US-0140473.
PR 22-OCT-1999; 99US-0140474.
PR 23-OCT-1999; 99US-0140475.
PR 24-OCT-1999; 99US-0140476.
PR 25-OCT-1999; 99US-0140477.
PR 26-OCT-1999; 99US-0140478.
PR 27-OCT-1999; 99US-0140479.
PR 28-OCT-1999; 99US-0140480.
PR 29-OCT-1999; 99US-0140481.
PR 30-OCT-1999; 99US-0140482.

Query Match 41.9%; Score 1684; DB 21; Length 775;
Best Local Similarity 44.8%; Pred. No. 1.3e-133;
Matches 354; Conservative 133; Mismatches 260; Indels 44; Gaps 15;


```
Db 527 ASDSRVFNIIISGTSMSCPHVSGLAALLKSVHPWSPAAIRSAALMTTAYKTYKDGKPLL 586
QY 597 D--GNKPAGVFAAGAHVNPQKAINPGLAVNIQPDVYITLCTLGFTSRDILAIHKVVS 654
Db 587 DIATGKPSPTFDGAGHVSPTTAINPGLIYDLTTEDYLGFLCALNYSPOIRSVRRNT 646
QY 655 CNGILRNKPGFS---LNPSIAVIFKRGKTTMTTRVTNVGSPNSIYSVNVKA-PEGIK 710
Db 647 CD----PSKSYSVADLNYSFAVND-GVCAKYKTRVTSVGAGTI-YGVKVTSETTGK 700
QY 711 VIVNPKRLVFKHVDQTLISRVWFVLKKNRGGKVAFAGQLTWNHNLMMORVSPISV 770
Db 701 ISVEPAVLNPEANEKKSYTVTFTVDSKPSG---SNSFGSIWSDGKHV---VGSVAI 754
QY 771 T 771
Db 755 S 755

RESULT 12
AAG43556
ID AAG43556 standard; Protein; 780 AA.
XX
AC AAG43556;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54455.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132483.
PR 14-MAY-1999; 99US-0132456.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
```

```
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 08-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 40.9%; Score 1642.5; DB 21; Length 780;
Matches 342; Conservative 128; Mismatches 281; Indels 33; Gaps 11;

Qy 4 KPFFLCI-----IFLLFCSSSEILOQTYIVQLHPNSETAKTFASKFDWHLSPLOEAVLG 59
Db 9 KPFLFILSINLIPLQAEITTOISTKTYVHM-DKSAMPLPYTNHLQWYSSKINSVTOH 67
Qy 60 VBEEEEPSSRLYSYSAIEGFAAQLTESEAEILRYSPEVAVRPHVLQVOTTSYK 119
Db 68 KSOEEGNRRILYTYQTAFHGLAAQLTOBEAERLEEDGWAVIPETRYELHTTSPTF 127
Qy 120 LGLDGFGNSGWSKSRFGQGTIIGVLDGTGWPPSPFDDTGMPISIRKWKIGIQESESFS 179
Db 128 LGLERQESERVAERVTDHVVVGVLDGTGWPPSPFDDTGMPISIRKWKIGIQESESFS 187
Qy 180 SSSCNRLIGARFFIRHVRVANSPEESPMPREVIASRDTSGHGTHTASTVGGSSYSMAN 239
Db 188 KRCNRKIVGARVYRGYEATGKIDE---ELYKPRDRDGHGTHTAATVAGSPVKGN 244
Qy 240 VLNGAGVARGMAPGAGHIAVYKVFNGCYSSDILAAIDVAIQDVVLSLSLGFPPIPL 299
Db 245 LFGFAYGTARGMAQKARVAAKYCVWGGCFSSDILSAVQAVADGVQVLSISLGGGVSTY 304
Qy 300 YDDTIAIGTFRAMERGISVCAAGNNGPTSSVANTAPWVSTIGAGTLDRRPPAVRLAN 359
Db 305 SRDLSIATTCAMEMGVFVSCSAGNGGPDPISTNVSPWITTVGASTMDRDPFATYKIGT 364
Qy 360 GKLLYGESLYPGKIKNAGREVEVIV---TGDKGSEPCLRGLSPREIRGKWKICDRG 416
Db 365 MRTFKGVSLYKGRVLPKNQYPLVILGRNASSPDPTSPCLDGLDRRHVAGKIVICDRG 424
Qy 417 VNGRSEKGEAVKAGGAGVAMILANTEINOEDSIDVHLLPATLICYTESVLLKAYVNATVK 476
Db 425 VTPRVQGVQVRAAGGIGMVLNTATNGEELVADSHMLPAVAVGEKGLIKQYAMTSKK 484
Qy 477 PKARIIFGGTVIGSRAPVAQFSARGPSIANPSILKPDMIAPGVNIIAAMPONLGTGL 536
Db 485 ATASLEILGTRIGIKPSPVAAFSRGNFSLSLKPDLLAPGVNIIAAMPONLGTGL 544
Qy 537 PYDSRVNFTVMSGTSNCPHVSIGITALIRSAVNPWSPAIKSALMTADLYDRQGKAIK 596
Db 545 SSDPRYKFNILSGTSNCPHVSIGVAAIKSRHPDWSPAIKSALMTATYVHDNMFPLT 604
Qy 597 D--GNKPAGVFAIGAGHVPKAINPGLVYNIQVDYITVLCITLGTFRSDILAIT-HKNV 653
Db 605 DASGAAPSSPDYDHGAGHIDFLRATDPLGVYDIGPQEIFELCTQDLSPSQLKVFTHSNR 664
Qy 654 SCNGILKRNQFSLNYPISIAVIFRKGKTTMIT--RRTVNVGSPNSISYVNVKAPGIVK 711
Db 665 TCKHTLAKNEG-NLNYPAISALPENTHVKAMTLRRVTVNVGPHISSYKVSVPFKGASV 723
Qy 712 IVNPKRLVFKHVDQTLISY----RWVFLVKKNRGGKVSAFQQLTWNHNLMOVRSP 767
Db 724 TVQPKTLNF--SKHOKLSYTVTFRPRMKRPEFG-----LVMKST---THKVRSP 771
Qy 768 ISVT 771
Db 772 VIIT 775

RESULT 13
ABB93785
ID ABB93785 standard; Protein; 780 AA.
AC ABB93785;
XX
XX
DT 31-MAY-2002 (first entry)
XX
```

```

DE XX Herbicidally active polypeptide SEQ ID NO 2996.
KW XX Herbicidal; plant; agriculture; herbicide.
XX XX
OS XX Arabidopsis thaliana.
XX XX
PN XX WO200210210-A2.
PD XX 07-FEB-2002.
XX XX
PF XX 28-AUG-2001; 2001WO-EP09892.
XX XX
PR XX 28-AUG-2001; 2001WO-EP09892.
XX XX
PA XX (FARB ) BAYER AG.
XX XX
PI XX Tietjen K, Weidler M;
XX XX
DR XX WPI; 2002-269010/31.
XX XX
PT XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX XX
PS Claim 5; SEQ ID NO 2996; 261pp + Sequence Listing; English.
XX XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX XX
SQ Sequence 780 AA;
Query Match 40.9%; Score 1642.5; DB 23; Length 780;
Best Local Similarity 43.6%; Pred. No. 4.4e-130;
Matches 342; Conservative 128; Mismatches 281; Indels 33; Gaps 11;
QY 4 KPFPCLT-----IFLFCSSSEILQQTIVQLHNSFETAKTPASKFDWHLFLOQAVLG 59
DB 9 KPFLFILSLINFLQAEITTTQISTKTYVIHM-DKSAMPLPYTNHLQWSSKINSVTQH 67
QY 60 VEREEEPSSRLYSYGSAEGFAAQLTESEAILAYSPEVWVRPDHVLQVQTTYSYKF 119
DB 68 KSOEEGNNRILTYQTAFHGLAAQLTQEEARLEEDGVVAIVETRYELHTRSPTF 127
QY 120 LGLDGFNGSVGWSKRFQGGTIIIGVLDTGVPSPSFDGTGMPISIPKKNKICQEGESTS 179
DB 128 LGLERQSERVWAERVTDHVVGVLDTGVPSESEFNDTGMPSPVATWRGACETGKRL 187
QY 180 SSSCNRLIGARFIRHRVANSPESPNMPREYISARSTGHTGHTASTVGGSSVSMAN 239
DB 188 KRCNCRKIVGARFVRYGAEATGKIDE---ELEVKSPDRDRGHTHTAATVAGSPYKGAN 244
QY 240 VLNGAGVARGMAPGAHIAKYKWCNFCYSSDILAAIDVAIQDKVDVLSLGGPPIPL 299
DB 245 LFGFAYTARGMAQKARVAAYKYKWCVGGCFSSDILSAVDQAVADGVQLSISLGGGVSTY 304
QY 300 YDDTIAIGTFRAMEGIVSICAGNNGPIESSVANTAPVSTIGACTLDRRFPVVRLAN 359
DB 305 SRDLSLATTGAMMGVVFVSCSAGNGGPDISTNVSPWITTVGASTMDRDPFATVIGT 364
QY 360 GKLLYGESLYPGKGIKNAGREVEYIV---TGDKGSEFCRLGSLPREIRGKMWICDRG 416
DB 365 MRITFKGSLYKGRVLPKNQYPLVYLGKRVNASSPDPTSFCLDGLDRHVGAKIVICDRG 424
QY 417 VNGRSEKGEAVKAGGVAMTILANTEINQEEEDSDVHLLPATLIGYTESVLLKAYNATVK 476

```

```

DB 425 VTPRVQKQVVVRAGGIGMVLNTATNGEELVADSHMLPAVAVGEGEKLKIQYAMTSKK 484
QY 477 PKARIIFGGTVIGSRAPAEVAQFSARGPSLANPSILKPDMLAPGVNIIIAAPQNLGPTGL 536
DB 485 ATASLEILGTRIGIKPSVVAASFSSRGPNFJLSLEILKPDMLAPGVNIIIAAWTGDMAFSSL 544
QY 537 PYDSRRVNFVTWMSGTSMSCPHVSGITALLIRGAYPNWSPAAIKSALMTTADLYDRQGAIK 596
DB 545 SSDPRVKFNILSGTSMSCPHVSGVAALIKSPHDPSPAAIKSALMTTAYVVDNMFELT 604
QY 597 D--GNKPAGVFAIGAGHVNPOKAINPGLVNIQVDYITYCTIGFTRSDILAIT-HQNV 653
DB 605 DASGAAPSSPYDHGAGHIDPLRATDPLGVYDIGQVEYFEFLCTQDLSPLKVTKEHSNR 664
QY 654 SCNGILRNKPGFSLNYPISIAVIFKRGKTEMIT--RRVTNVGSPNSIYVNVKAPGKIKV 711
DB 665 TCKETLAKNPG-NLNYPAISALFPENTHVKAMTLRRTVTNVGPHISSYKVSVPFKGASV 723
QY 712 IVPKRLVFKHVDOTLSY---RVYFVLKKNRGKGVASFAQGOLTWVNSHNLQVRVSP 767
DB 724 TVQPKTLNFTSKHQKLSYTVTRFRFMKRPFEGG-----LVWKST---THKVRSP 771
QY 768 ISVT 771
DB 772 VIIT 775
XX XX
RESULT 14
AAG43558
ID AAG43558 standard; Protein; 735 AA.
XX XX
AC AAG43558;
XX XX
DT 18-OCT-2000 (first entry)
XX XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54457.
XX XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX XX
OS Arabidopsis thaliana.
XX XX
PN EPI033405-A2.
XX XX
PD 06-SEP-2000.
XX XX
PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.

```

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145916.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.


```

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-01511303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 40.1%; Score 1612; DB 21; Length 740;
Best Local Similarity 44.7%; Pred. No. 1.6e-127;
Matches 332; Conservative 119; Mismatches 263; Indels 28; Gaps 9;

Qy 42 FASKFDHLSFLOEAVLGVEEEFSSRLLSYSGAIFGAQLTSEAEILRYSPEV 101
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 10 YTNHLQWYSSKINSVTQHKSQEEGNRRILYTYQTAHFHGLAAQLTQEEAELEEDGV 69
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 102 AVRPDHLVQVQTTYSYKFLGLDGFNGSVKSRFCQGTIIIVLDTGVWPESPSDDTGM 161
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 AVIPETRYELHTTRSTFLGLEQESERVWAERVTDHVDVVGVLDTGIWPESESFNDTGM 129
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 162 PSIPRWKIGICQEGESFSSSSCNKRKLGARFFIRHRVANSPEESNMPREYISARDSTG 221
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 SPVPATWRGACETGKFLKNCNKRKIVGARVYRGYEATGKIDE---ELEYKSPDRUG 186
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 222 HGHTASTYGGSSVSMANVLGNAGVARGAPCAHIAIVYKVCWFNGCYSDILAAIDVAI 281
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 HGHTAATVAGSPVKGANLFGPAYGTARGWAQKARVAAYKVCWVGCGCFSSDILSAVDQAV 246
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 282 QDKVDVLSLGGFPDPLDYDDTIAITGTFRAMEGSGIVCAAGNNGPIESSVANTAPWVST 341
: : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 247 ADGVQVLSISLGGGVSTYRDSLSIATFGAMEMGVFVSCSAGNCGPDPISLUTNVSPWITT 306
QY 342 IGAGTILDRFPFAVVRILANGKLYGSLYPKGIKNAGREVEYIV---TGGDKGSEFCLR 398
Db 307 VGASTWDRDFPAIVTKGTWTKGVSLYKGRIVLPKNQYPLVYLGRNASSPDPTSPCLD 366
QY 399 GSLPREIEIKGMVICDRGVNGRSEKEAVEKEAGGVAMILANTEINOEDSDIVHLLPATL 458
Db 367 GALDRRHVAGKIVICDRGVTVPVQKGVVVRAGGIGWLTNTATNGEELVADSHMLPAVA 426
QY 459 IGYTESVLLKAYNAIVKPKARIIFGTVIGRSRAPEVAQPSARGPSLANPSILKPDMLA 518
Db 427 VGEKEGLIKQYAMTSKKATASLEILGTIGIKPSPFWAAAFSSRGPNFELSUEILKPDLLA 486
QY 519 PGVNIIAWPNQGLPTGLPYDGRNVFTVMSGTSMSCPHVSGITAIIRSAVENWSPAAIK 578
Db 487 PGVNIILAAWTGDWAPSSLSDDPRVVKFNILSGTSMSCPHVSGVAALIKSRHPDWSPPAAIK 546
QY 579 SALMTTADLYDRQKAID--GNKPAGVPAIGAGHVNPOKAINPGLVYNIQVDYITYLC 636
Db 547 SALMTTAYVHDNMFKELTDASGAAPSSPYDHGAGHIDPLRATDPGLVYDIGPQEVFEFLC 606
QY 637 TLGFTRSDLAIT-HQVNSCNGILRNKPGFSLNYPISIAVIFKRGKTTMIT--RRVTNVG 693
Db 607 TODLSPSQLKVETKHSNRTCKHTLAKNPG-NLNYPAISALFPFENTHVKAMTLRRTVTNVG 665
QY 694 SPNSIYSVNVKAPGKIVNPKRLVFKHVDQTLSY---RVWFVLKKKNGGKVASFQAQ 749
Db 666 PHISSYKVSVPFGASVTVPKTNFTSKHOKLSYTVTRFRFRKRPEFGG----- 718
QY 750 GOLTWVNSHMLQORVSPISVT 771
Db 719 --LVWKST---THKVSPIIIT 735

Search completed: February 2, 2004, 09:10:01
Job time : 49 secs

US-08-894-818B-5

Query Match 8.6%; Score 345; DB 3; Length 659;
Best Local Similarity 23.2%; Pred. No. 3.2e-22;
Matches 145; Conservative 62; Mismatches 145; Indels 272; Gaps 21;

QY 29 IVQLHPNSETAKTFASKFDWHLFLQEAFLGVVEEPESSRLLYSYGS-----AIEG 81
DB 50 IQKLNPEEISTVIV--FENH-----REKEIAVRVLELMGAKVRYVYHIIPA 94

QY 82 FAQA-----LTESAEI--LRYSPVAVR-----PDVLOVQVITYSKFL 120
DB 95 IAADLKVRDLVIVSLTGGKAKLSGVRFIOEDYKVTVSAELEGDESAAQVMATY----- 149

QY 121 GLDGFNGSVNKSFRQGTIIIGVLTGVNWPESPSDDTCMPSPRKKWGIQCEGESFSS 180
DB 150 -----VNNLYGDSGGTIGITGI-----DASHDDL----- 176

QY 181 SSCNRKLIGARFFIRGHRVANSPEESPMPREYISARDSTGHGTHPTASTVGGSSVMANV 240
DB 177 ---QGVIGWVDFUNG-----RSY--PYDDHGHGTHVASIAAGTGAA--- 213

QY 241 LMGAGVARGMAFGAHIAVYKVCWFNGCYS--SDILAAIDVAIQDK-----VDVLSLSLGGF 295
DB 214 ---SNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYGIKVINLSLSS 270

QY 296 PIPLYDDTIALGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTLDRRFPVAV 355
DB 271 QSSDGTDSLQAVNNAWDAGIVCVAAAGNPGNTYTVGSPAASKVITVGAVDN----- 325

QY 356 RLANKLLYGESLYPGKIKNAGREVEVIVVTGDKGSEFCLRGSLPREEIRKGVICDR 415
DB 326 ----- 325

QY 416 GVNGRSEKGEAVKAGGAVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNATV 475
DB 326 ----- 325

QY 476 KPKARIIEGTVIGRSRAPEVAQFSARGPSLANPSILKPDMTAPGVNIIAAMPQNLGPT- 534
DB 326 -----DNIAFSRGPPTA--DGRLKPEVWAPGVVDIIA--PRASGTSM 363

QY 535 GLPYDSRRVNTVMGTSMSCPHVGITALIRSAYPNWSAAIKSALMTTADLYDRQOKA 594
DB 364 GTPIND---YTTAAGTSMATPHVGVGALILQAHPSWTPDKVKTLIETADIV---APKE 418

QY 595 IKGNKPEAGVFAICAGHVPQKAI 618
DB 419 IAD-----IAYGAGRVNYYKAI 435

RESULT 2

US-08-894-818B-35
Sequence 35, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/894,818B
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-894-818B-35

Query Match 8.3%; Score 332; DB 3; Length 654;
Best Local Similarity 22.4%; Pred. No. 4.7e-21;
Matches 140; Conservative 65; Mismatches 147; Indels 274; Gaps 21;

QY 29 IVQLHPNSETAKTFASKFDWHLFLQEAFLGVVEEPESSRLLYSYGS-----AIEG 81
DB 50 IQKLNPEEISTVIV--FENH-----REKEIAVRVLELMGAKVRYVYHIIPA 94

QY 82 FAQA-----LTESAEI--LRYSPVAVR-----PDVLOVQVITYSKFL 120
DB 95 IAADLKVRDLVIVSLTGGKAKLSGVRFIOEDYKVTVSAELEGDESAAQVMATY----- 149

QY 121 GLDGFNGSVNKSFRQGTIIIGVLTGVNWPESPSDDTCMPSPRKKWGIQCEGESFSS 180
DB 150 -----VNNLYGDSGGTIGITGI-----DASHDDL----- 176

QY 181 SSCNRKLIGARFFIRGHRVANSPEESPMPREYISARDSTGHGTHPTASTVGGSSVMANV 240
DB 177 ---QGVIGWVDFUNG-----RSY--PYDDHGHGTHVASIAAGTGAA--- 213

QY 241 LMGAGVARGMAFGAHIAVYKVCWFNGCYS--SDILAAIDVAIQDK-----VDVLSLSLGGF 295
DB 214 ---SNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYGIKVINLSLSS 270

QY 296 PIPLYDDTIALGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTLDRRFPVAV 355
DB 271 QSSDGTDSLQAVNNAWDAGLVVVAAGNPGNPKYITIGSPAAASKVITVGAVDK----- 324

QY 356 RLANKLLYGESLYPGKIKNAGREVEVIVVTGDKGSEFCLRGSLPREEIRKGVICDR 415
DB 325 ----- 324

QY 416 GVNGRSEKGEAVKAGGAVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNATV 475
DB 325 -----YDV----- 327

QY 476 KPKARIIEGTVIGRSRAPEVAQFSARGPSLANPSILKPDMTAPGVNIIAAMPQNLGPTG 535
DB 328 -----ITSFSSRGPTA--DGRLKPEVWAPGVNIIAARASGTS--WG 364

QY 536 LPYDSRRVNTVMGTSMSCPHVGITALIRSAYPNWSAAIKSALMTTADLYDRQOKAI 595
DB 365 QPIND---YTTAAGTSMATPHVGVGALILQAHPSWTPDKVKTLIETADIV----- 414

QY 596 KGNKPGV--FAIGAGHVNPOKAIN 619
DB 415 ----KPDEIADIAYGAGRVNAYKAIN 436

RESULT 3

US-09-445-472-16
; Sequence 16, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 16
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-445-472-16

Query Match 8.3%; Score 332; DB 4; Length 654;

Best Local Similarity 22.4%; Pred. No. 4.7e-21;

Matches 140; Conservative 65; Mismatches 147; Indels 274; Gaps 21;

QY 29 IVOLHNPSETAKTFASKFDWHLFLQEAVLGVEEEREPSSRLLYSGS-----AEG 81
DB 50 IQKLNPEESTIV--FENH-----REKEIAVRLEMLGAKRVYVHIIPA 94
QY 82 FAAQ-----LTESEAEI--LRYSPVAVR-----PDHVLQVQTTYSYKFL 120
DB 95 IADLKVRDLVLSGTTGGKAKUSGVRFQEDYKVTVAELGLEDSAAQVWATY----- 149
QY 121 GLDGFNGSVWKSRSRFGQGTIGVLDTGVPWSPSPDDTGMPSPRKNKRGICQEGESFSS 180
DB 150 -----VNNLYGDSGTIGIITGI-----DASHEDL----- 176
QY 181 SSCNRKLIARFIRHRVANSPEESPMPREVISARDSTGHTTASTVGGSSVSMANV 240
DB 177 ---QGVIGWVDFVNG-----RSY--PYDDHGHTHVASTAAGTGAA--- 213
QY 241 LGNGAGVARGMAGAHIAVYKVCWFNGCYS--SDILAAIDVAIQDK-----VDVLSLSLGGF 295
DB 214 ---SNGKYGMAPCAKLAGIKVLGADGSGSISTIIKGVAVDNKDKYGIKVINLSGSS 270
QY 296 PIPLYDDTIAIGTFRAMEGSGISVICAAGNPGTRESSVANTAPWSTIGAGTIDRRFPVAV 355
DB 271 QSSDGTALSOAVNAWADAGLVVVAAGSGGNKNTYIGSPAASKVITVGAVDK----- 324
QY 356 RLANKLLYGESLYPGKIGKNAGREVEVIYVTGGDXGBFLRGLSPREEIRGKMWICDR 415
DB 325 -----YDV----- 324
QY 416 GVNGRSEKBAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYNATV 475
DB 325 -----YDV----- 327
QY 476 KPKARIIFGCTVIGRSRAPEVAQFSARGPSLANPSTLKPDMIAPGVNIITAAWPNGLPTG 535
DB 328 -----ITSFSRGPPTA---DGLKPEVAVAPGNWIIAASAGTS-MG 364
QY 536 LPYDSRRVNTVMSTGSMCPHYSGITALIRSAYPNWSAAIKSALMTTADLYDRQKAI 595
DB 365 QPIND---YTAAPGTSMATPHVAGIAALLQAPSWTPDKVKYKTALETADIV----- 414

QY 596 KGNKPGV--FAIGAGHVNPOKAIN 619
DB 415 ----KPDEIADIAYGAGRVNAYKAIN 436

RESULT 4

US-08-894-818B-1
; Sequence 1, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 659 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-894-818B-1

Query Match 8.2%; Score 330; DB 3; Length 659;

Best Local Similarity 23.7%; Pred. No. 7.2e-21;

Matches 124; Conservative 57; Mismatches 119; Indels 224; Gaps 17;

QY 103 VRPHVLQVQTTYSYKFLGLDGFNGSVWKSRSF-GQGTIGVLDTGVPWSPSFDDTGM 161
DB 128 IQEDYKQVQDDATSVSQIGAD-----TVWNSLGYDGGVVVAIVDTGI-----DANH 174
QY 162 PSIRPKKKGICQEGESFSSSCNCKLIGARFFIRHRVANSPEESPMPREVISARDSTG 221
DB 175 PDL-----KG-----KVLGWTDAVNGRS-----TPYDDQG 199
QY 222 HGHTTASTVGGG--SVSMANVLGNCAGVARGMAGAHIAVYKVCWFNGCYS--SDILAAIDV 279
DB 200 HGHTVAGIVAGTGSVNQYI-----GVAFGAKLVGVKVLGADGSGSVSTIIAGVDW 250

QY 525 AAMPQNLGPTGLPYDSRRVNFVMSGTSMSCHVSGITALIRSAVNPNSPAAIKSALMTT 584
Db 223 AARASGTS-MGQPIN--YTTAAPTGMTATPHVAGIAALLLOAHPSWTPDKVKTALET 278
QY 585 ADLYDRQKAIKDGKPAV--FAIGAGHVNPOKAIN 619
Db 279 ADIV-----KPDEIADIYAGGRVNAKAIN 304

RESULT 7
US-08-894-818B-3
; Sequence 3, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:

; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: /note= Xaa at position 428 is Gly or Val.

US-08-894-818B-3

Query Match 8.1%; Score 324.5; DB 3; Length 522;
Best Local Similarity 23.0%; Pred. No. 1.5e-20;
Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;

QY 110 QVOTTSYKFLGLDGFNGSVKSRFGQGTIGVLDTGWVPESFDDTGMPSPRKWK 169
Db 12 QVWATY-----VWNLGYDGGSGITIGIDTGI-----DASHPDL----- 44
QY 170 GICQEGESFSSSSCNKLCARFIRGHRVANSPEPSNMPREYISARDSTGHGHTAST 229

Db 45 -----QKVGWVDFVNG-----RSY--PYDDHGHGTHVASI 74
QY 230 VGGSSVSMANVLGNAGVARGMAPGAHIAVYKVCWFGCYS--SDILAAIDVAIQDK--- 284
Db 75 AAGTGAA-----SNGKYKGMAPGAKLAGIKVLGADGSGSIITIKGVWAVDNKDKYG 127
QY 285 VDVLSLSLGGFPPIPLYDDTIAICTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGA 344
Db 128 IKVINLSLSSQSSDGTALSOAVNAADAGLVVVVAAGNSGPNKYTIGSPAAASKVITV 187
QY 345 GTLDRRPPAVRLANGKLLYGESLYPGKIGIKNAGREVEVIYVTGGDKGSFCLRGLSPRE 404
Db 188 GAVDK----- 192
QY 405 EIRGKMWICDRGVNRSEKGEAVKEAGVAMILANTEINOEDSIDVHLLPATLIGYTES 464
Db 193 -----YDV----- 195
QY 465 VLLKAYVNAVTKPKARIIFGGTVIGRSRAPEVAQFARGPSLANPSILKPDMTIAPGVNII 524
Db 196 -----ITSFSSRGPTA--DGRLLKEVVAVPAGNII 222
QY 525 AAMPQNLGPTGLPYDSRRVNFVMSGTSMSCHVSGITALIRSAVNPNSPAAIKSALMTT 584
Db 223 AARASGTS-MGQPIN--YTTAAPTGMTATPHVAGIAALLLOAHPSWTPDKVKTALET 278
QY 585 ADLYDRQKAIKDGKPAV--FAIGAGHVNPOKAIN 619
Db 279 ADIV-----KPDEIADIYAGGRVNAKAIN 304

RESULT 8
US-09-445-472-4
; Sequence 4, Application US/09445472
; Patent No. 5358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (428)..(428)
; OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-09-445-472-4

Query Match 8.1%; Score 324.5; DB 4; Length 522;
Best Local Similarity 23.0%; Pred. No. 1.5e-20;
Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;

QY 110 QVOTTSYKFLGLDGFNGSVKSRFGQGTIGVLDTGWVPESFDDTGMPSPRKWK 169
Db 12 QVWATY-----VWNLGYDGGSGITIGIDTGI-----DASHPDL----- 44
QY 170 GICQEGESFSSSSCNKLCARFIRGHRVANSPEPSNMPREYISARDSTGHGHTAST 229
Db 45 -----QKVGWVDFVNG-----RSY--PYDDHGHGTHVASI 74
QY 230 VGGSSVSMANVLGNAGVARGMAPGAHIAVYKVCWFGCYS--SDILAAIDVAIQDK--- 284

Db 75 AAGTGAA-----SNGKYKGNAPAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYG 127
Qy 285 VDVLSLSLGPPPIPLYDDTIAIGTFRAMEGIVSICAAAGNGPIESSVANTAPWVSTIGA 344
Db 128 IKVINLSGSSOSDGTALSOAVNAWDAGLVVVAAGNSGPKYIIGSPAAASKVITV 187
Qy 345 GTLDRFPFAVRLANGKLLYGBSLYPGKGIKNAGREVEIYVTGDKGSEFCRLGSLPRE 404
Db 188 GAVDK-----YDV----- 192
Qy 405 EIRGKVICDRGVRSEKGEAVKEAGGVAMILANTEINQEEDSIDVHLLPATLIGYTES 464
Db 193 -----YDV----- 195
Qy 465 VLLKAVVNAVTKPARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDWIAPGVNI 524
Db 196 -----ITSFSSRGPTA--DGRUKPEVAVPGVNI 222
Qy 525 AAWPQNLGTLGPLYDSRRVNFVTMSGCPHVSIGITALIRSAYPNWSPPAAIKSALMTT 584
Db 223 AARASGTS--MGQIPND---YVTAAGTSMATPHVAGIAALLLQAHPSWTPDKVXTALLET 278
Qy 585 ADLYDRQKAIKGNKPNAGV--FAIGAGHVNPQKAIN 619
Db 279 ADIV-----KPDEIADIAGAGRVNAVYKAIN 304

RESULT 9

US-09-000-016-4
; Sequence 4, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 734 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-000-016-4

Query Match 7.4%; Score 297; DB 3; Length 734;
Best Local Similarity 21.4%; Pred. No. 8.2e-18;
Matches 155; Conservative 70; Mismatches 224; Indels 276; Gaps 22;
Qy 36 SETAKTFAS-KPDWHLFLQEAFLGVVEEBEPSSRLLYSGSAIEGFAAQ----- 85
Db 99 AARLVASGKLDRLFDITELGKAATRNQSQGLKVIIVGYQGAARAAKAEVREAGELR 158
Qy 86 -LTSEAEILRYSPEVAVRPHVLQVQVITYS-YKFLGLDG-----FNGSGV 131
Db 159 TLTSLNADAVRPHEDASELWDVINGDRTASGIAHVLWDGVVRAALDTSVQIGAPKAW 218
Qy 132 SKSRFOGQIIIGVLTGVPWSPSPDDTGMPSPKPKWKICQEGESFSSSSCNRLKILGAR 191
Db 219 SAGYDCKGVKIAVLDTGV-----DTSHPDL-----KGRVTASKNFTAAP----- 257
Qy 192 FFIRGHRVANSPEESPMPREVIASARDSTGHCHTHTASTVCGSSVSMANVLNGAGVARGM 251
Db 258 -----GAGDKVGHGTHVASIAAGTG-----AQSGKYKGV 287
Qy 252 APGAHIYVVKWFGNC-YSSDILAAIDVAIQKVDVLSLSLGGFPPIPLYDDTIAIGTFR 310
Db 288 APGAALNGKVLDDSGFGDDSGILAGMEWAAAGQADVVNMSLGGMDTETDPLEAAVDKL 347
Qy 311 AMERGIVTCAAGNNGPIESSVANTAPWVSTICAGTLDRRFPVAVVRLANGKLLYGESLYP 370
Db 348 SAEKGVLFALAGNEGPEISGSPGSADAALTVGA----- 381
Qy 371 GKGIKNAGREVEIYVTGDKGSEFCRLGSLPREIRGKMWICDRGVNGRSEKGEAVKEA 430
Db 382 -----VDDKDK----- 387
Qy 431 GGVAMILANTEINQEEDSIDVHLLPATLIGYTESVLLKAVVNAVTKPARIIFGGTVIGR 490
Db 388 ----- 387
Qy 491 SRAPEVAQFSARGPSLANPSILKPDWIAPGVNIIAWPO-----NLGPTGLPYDSRR 542
Db 388 -----LADFSSTGPRLGDCAL-KPDTAFGVDITAASAEAGNDIGQEVGSGPAG----- 434
Qy 543 VNFTVMSGTSMSCPHVSIGITALIRSAYPNWSPPAAIKSALMTTADLYDRGKAIKGNKPA 602
Db 435 --YMTISGTSMATPHVAGAAALLKQHPDWTSAELKGALE-----TGSTKG 477
Qy 603 G---VFAIGAGHVNPQKAINPGLV-----YNIQ-----PV-DIYILCTLIGTFS 643
Db 478 GKYPTEQSGSGRIQADKALQQTVIADPVSFVGQWQPHETDDEPVTQKLYR-NLG-TQD 535
Qy 644 DILAITHKNVSCNGILRKNPGPSLNYPSIAVIFKRGKTTETLRRVTVNVSNSIYSVNV 703
Db 536 VTUKLTSTATDPKGAAPAGFTLGAITTVTVFAGGSASVDM-TADTRLGCTVDGAYSATV 594
Qy 704 KAPEG 708
Db 595 VATGG 599

RESULT 10

US-09-514-340-4
; Sequence 4, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE D
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.

DB 435 --YMTISGTSMTATPHVAGAAALRQHPDWTSAELKGAAL-----TGSTKG 477
QY 603 G---VFAIGAGHVNPQKAINPLV-----YNIQ-----PV-DVITVLTCLGFTFRS 643
DB 478 KYTFPEQSGRIQADKALQOTVIADPVSVSGVQWPHHTDDEPTKQITTR-NLG-TQD 535
QY 644 DILAITHKRVSCNGILRKNPGFSLNYPISIAVIFRGRKTKTEMITRRVTNVGSPNSIYSVNV 703
DB 536 VTLKLTSTATDPKGAAPAGFTLGAITVTVPAAGGSASVDM-TADTRLGCTVDGAYSAYV 594
QY 704 KAPEG 708
DB 595 VATGG 599
RESULT 11
US-09-000-016-2
; Sequence 2, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERI
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 823 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-000-016-2
Query Match 7.4%; Score 297; DB 3; Length 823;
Best Local Similarity 21.4%; Pred. No. 9.9e-18;
Matches 155; Conservative 70; Mismatches 224; Indels 276; Gaps 22;
QY 36 SETAKTPAS-KFDWHLSTFLOEAVLGVVEEESRLLSYSGAIEGFAAQ-----85
DB 99 ADAARLVASGKLDRLRFDITELGKAATRNQKGLKVIYGYQGAARAAKAEVREAGELRR 158
QY 86 -LTSSEAEILRYSPVAVRPHVLQVQTTYS-YKFLGLDG-----FGNSGVW 131
DB 159 TLTSLNADAVRTHEDASELWDVNTGDRGTASGIAHVWLDGVRRALDTSVQIGAPKAW 218

COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 734 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-514-340-4
Query Match 7.4%; Score 297; DB 4; Length 734;
Best Local Similarity 21.4%; Pred. No. 8.2e-18;
Matches 155; Conservative 70; Mismatches 224; Indels 276; Gaps 22;
QY 36 SETAKTPAS-KFDWHLSTFLOEAVLGVVEEESRLLSYSGAIEGFAAQ-----85
DB 99 ADAARLVASGKLDRLRFDITELGKAATRNQKGLKVIYGYQGAARAAKAEVREAGELRR 158
QY 86 -LTSSEAEILRYSPVAVRPHVLQVQTTYS-YKFLGLDG-----FGNSGVW 131
DB 159 TLTSLNADAVRTHEDASELWDVNTGDRGTASGIAHVWLDGVRRALDTSVQIGAPKAW 218
QY 132 SKSRFGGTTIGVLDGFWPESFDDTGMPSPTRKWKICQEGESFSSSSCNKRLIGAR 191
DB 219 SAGYDGGKGVIAVLDTG-----DTSHPDJ-----KCRVTSKNTAAP-----257
QY 192 FFRGRHVRANSPEESPMPREYISARDSTGHGHTASTVGGSSVSMANVLNGAGVARGM 251
DB 258 -----GAGDKVGHGTHVASIAAGTG-----AOSKGYKEV 287
QY 252 APGAHIAYKVCWFNGC-YSSDILAAIDVAIQKVDVLSLSLGGFFPIPLYDDTIAIGTFR 310
DB 288 APGAAILNGKVLDSGDDSGILAGMEWAAAGADVNMVSLGMDTPTDPLEAAVDKL 347
QY 311 AMERGSIVCAAGNNGPIESSVANTAPWSTTGAGTLDRRFPVAVRLANGKLLYGESLYP 370
DB 348 SAEGKVLFAAAGNEGPESTGSGSADAALTGA-----381
QY 371 GKGIKNAGREVEYIYTGDKGSEFCIRGSLFREEIRKMWICDRGVNRSEKGEAVKEA 430
DB 382 -----VDDKDK-----387
QY 431 GGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAVVNAVTKPKARIIFGGTVIGR 490
DB 388 -----387
QY 491 SRAPEVAQFSARGFSLANPSILKPDMLAPGVNIIAAWPQ-----NLGPTGLPYDSRR 542
DB 388 -----LADFSTGTGRLGDGAL-KPDVTAPGVDDITAASAEGNDIGQEVGGGPAG-----434
QY 543 VNTVWSGTSMSCPHVSGITALIRSAYPNWSPAAIKSALMTTADLYDRGKAIKGNKPA 602

DB 435 --YMTISGTSMTATPHVAGAAALRQHPDWTSAELKGAAL-----TGSTKG 477
QY 603 G---VFAIGAGHVNPQKAINPLV-----YNIQ-----PV-DVITVLTCLGFTFRS 643
DB 478 KYTFPEQSGRIQADKALQOTVIADPVSVSGVQWPHHTDDEPTKQITTR-NLG-TQD 535
QY 644 DILAITHKRVSCNGILRKNPGFSLNYPISIAVIFRGRKTKTEMITRRVTNVGSPNSIYSVNV 703
DB 536 VTLKLTSTATDPKGAAPAGFTLGAITVTVPAAGGSASVDM-TADTRLGCTVDGAYSAYV 594
QY 704 KAPEG 708
DB 595 VATGG 599
RESULT 11
US-09-000-016-2
; Sequence 2, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERI
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 823 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-000-016-2
Query Match 7.4%; Score 297; DB 3; Length 823;
Best Local Similarity 21.4%; Pred. No. 9.9e-18;
Matches 155; Conservative 70; Mismatches 224; Indels 276; Gaps 22;
QY 36 SETAKTPAS-KFDWHLSTFLOEAVLGVVEEESRLLSYSGAIEGFAAQ-----85
DB 99 ADAARLVASGKLDRLRFDITELGKAATRNQKGLKVIYGYQGAARAAKAEVREAGELRR 158
QY 86 -LTSSEAEILRYSPVAVRPHVLQVQTTYS-YKFLGLDG-----FGNSGVW 131
DB 159 TLTSLNADAVRTHEDASELWDVNTGDRGTASGIAHVWLDGVRRALDTSVQIGAPKAW 218

Db 595 VATGG 599

RESULT 13

US-09-000-016-7

; Sequence 7, Application US/09000016

; Patent No. 6143541

; GENERAL INFORMATION:

; APPLICANT: Akira ARISAWA et al.

; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV

; TITLE OF INVENTION: ITS EXPRESSION PRODUCT

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

; STREET: 2033 K Street, N.W., #800

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20006

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/000,016

; FILING DATE: January 30, 1998

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Warren M. Cheek, Jr.

; REGISTRATION NUMBER: 33,367

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-721-8200

; TELEFAX: 202-721-8250

; TELEX:

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 520 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-000-016-7

Query Match 7.0%; Score 283; DB 3; Length 520;

Best Local Similarity 21.8%; Pred. No. 8.5e-17;

Matches 133; Conservative 56; Mismatches 170; Indels 252; Gaps 18;

QY 126 GNSGVWKSRRPGQGTIGVLDTGVWVPSPSFDDTGMPISIPRKWKGICQEGESFSSSCNR 185

DB 9 GAPKWSAGYDGKGVKIAVLDTGV-----DTSHPDL-----KGRVTASKNFTAAP--- 53

QY 186 KLIGATFIRHVRVANSPEESPNNPREVISARDSTGHGTHTASTVGGSSVSMANVLNGA 245

DB 54 -----GAGDKVGHGTHVASIAAGTG-----AQSK 77

QY 246 GVAGMAPGAHIAVYKVCWFNGC-YSSDILAAIDVAIQDKVDVLSLSLGGPPIPLYDDTI 304

DB 78 KKYKGVAPGAAILNGKVLDDSGFGDDSGILAGMEWAAQAQADVVMSELGGMDTETDPLE 137

QY 305 AIGTFRAMERGISVTCAGNNGPIESSVANTAPVWSTIGACTLDRRFPVAVRLANGKLLY 364

DB 138 AAVDKLSAEKGVLPALRAAGNEGPIESGPSGADAALTVGA----- 177

QY 365 GESLYPGKIKVAGREVEVIYVTGDDKSGSEFCRLGSLPREIRKMWICDRGVNGRSEK 424

DB 178 -----VDDKDK----- 183

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:11:33 ; Search time 38 Seconds
(without alignments)

4238.517 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEKPPFLCIIIFLLFCSSS.....NSHNLQVRVPSIVLTKN 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1175	29.2	734	12	US-10-259-165-272
2	332	8.3	654	14	US-10-090-624-16
3	330	8.2	659	14	US-10-090-624-12
4	324.5	8.1	412	14	US-10-090-624-1
5	324.5	8.1	522	14	US-10-090-624-4
6	318.5	7.9	1208	15	US-10-156-761-13251
7	307	7.6	1139	15	US-10-156-761-10856
8	282	7.0	1079	15	US-10-112-488-39
9	276.5	6.9	382	11	US-09-813-408-7
10	271.5	6.8	382	12	US-10-146-905A-8
11	271.5	6.8	382	14	US-10-090-624-31
12	271.5	6.8	382	15	US-10-104-693-2
13	269.5	6.7	1150	10	US-09-870-122-3
14	266.5	6.6	382	12	US-10-146-905A-11
15	264.5	6.6	382	11	US-09-813-408-20

16	262	6.5	1237	12	US-10-314-657-4
17	261.5	6.5	1181	10	US-09-870-122-23
18	260.5	6.5	382	12	US-10-423-649-2
19	260.5	6.5	382	15	US-10-033-325-2
20	260.5	6.5	382	15	US-10-228-572-2
21	259	6.4	418	10	US-09-966-921A-2
22	257.5	6.4	380	11	US-09-813-408-19
23	257.5	6.4	1167	10	US-09-870-122-2
24	255.5	6.4	361	12	US-10-007-389-1
25	251.5	6.3	377	11	US-09-813-408-1
26	251	6.2	380	11	US-09-824-893A-261
27	248	6.2	380	12	US-10-324-152-8
28	248	6.2	380	12	US-10-324-152-9
29	248	6.2	381	10	US-09-920-118-16
30	247.5	6.2	357	9	US-09-837-235-15
31	245	6.1	375	11	US-09-813-408-8
32	244.5	6.1	379	11	US-09-813-408-11
33	243	6.0	269	11	US-09-779-334A-7
34	242.5	6.0	379	10	US-09-920-118-14
35	241	6.0	627	11	US-09-927-827-60
36	240	6.0	275	15	US-10-104-693-3
37	240	6.0	279	11	US-09-813-408-21
38	239	5.9	275	11	US-09-813-408-23
39	239	5.9	275	12	US-10-324-152-1
40	237.5	5.9	274	12	US-10-336-324-5
41	237.5	5.9	379	11	US-09-813-408-10
42	237.5	5.9	379	12	US-10-146-905A-10
43	237.5	5.9	379	15	US-10-202-339-2
44	237	5.9	275	8	US-08-322-678-7
45	237	5.9	275	9	US-09-060-854B-3

ALIGNMENTS

RESULT 1

US-10-259-165-272

; Sequence 272, Application US/10259165

; Publication No. US20030135888A1

; GENERAL INFORMATION:

; APPLICANT: Zhu, Tong

; APPLICANT: Wang, Xun

; APPLICANT: Chang, Hur-song

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Moughamer, Todd

; APPLICANT: Provart, Nicholas

; APPLICANT: Rieke, Darrell

; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING

; FILE REFERENCE: 70030-NP

; CURRENT APPLICATION NUMBER: US/10/259,165

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04

; PRIOR APPLICATION NUMBER: US 60/368,327

; PRIOR FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26

; NUMBER OF SEQ ID NOS: 782

; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta

; SEQ ID NO 272

; LENGTH: 734

; TYPE: PRT

; ORGANISM: Oryza sativa

; US-10-259-165-272

Query Match 29.2%; Score 1175; DB 12; Length 734;

Best Local Similarity 37.1%; Pred. No. 1.1e-102;

Matches 293; Conservative 121; Mismatches 249; Indels 126; Gaps 27;

QY 38 TAKTASKEFDEL-SFLOEAVLGVEEBEERS-----SRLLYSY 75
Db 4 TLSSPHNVVDWITTSFLSYIVTMGKXKDDPVTASHHDILTSLVSGKDGAMKSIYVSY 63
QY 76 GSAIBGFAQLTESEAEILRYSPEVAVRPDRHLVQVOTYSYKFLGLDGFNGSGWKSXR 135
Db 64 KHGFSGFAAMLTESQAEELAR-PEVISVKPNYHQATRTSNDWFLGLNYSQSLKCKAK 123
QY 136 FCGITIGVLDGWPESPDFTGMPISIPRKWGI COEGESFSSS--CNRLKGARFFI 194
Db 124 NGEDIVIGVIDSGIWPESRSPDNGYSPPVPAWKGCQTGAFAFNATTCNKKIIGVRWYS 183
QY 195 RGHVANSPEEPSPNPREVISARDSTGCTHTASTVGGSSVSMANVLNG--AGVARGM 251
Db 184 GG-----IPDE--NKGETYSARDJGGHGTAVSTIVGGQVNVNHRGCGMAGATARGG 236
QY 252 AFGAHVAVYKVCWFNG-----CYSSDILAAIDVAIOKVDVLSLSLGGFPPIPLYDDTTAIG 307
Db 237 APPARVAVYKVCW--GLRAQCGGAALAAIDAMNDGVDVLSLSIGG-----AG 283
QY 308 ----TFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGATLDRRPPAVVRLANGKLL 363
Db 284 EHYETLHVARGIPVVGNDGPTPQIVRNTVPVWITVAASTIDRAPFTVISLGNKKF 343
QY 364 YGESLYPGKGIKNAGREVVIYVGGDKGSEBCLRGSLPREIRCKWICD-----414
Db 344 VQOSLYNATASSTKQMLV-----DGSS--CDTQTLASINITSKVLCSPSPSILMPRL 395
QY 415 --RGVNGRSEKGEAVKEAGGVAMILANTEINOEBSDIDV---HLLPATLIGYTESVLLKA 469
Db 396 SLGDIIGR-----VTKAGANGLI FVOYSVSNALDFLNACRASVPCVLVDYEITRRIES 449
QY 470 YVNATVPKARIIFGCTVIGRS--RAPEVAQFSARGPSLANPSILKPDMLAPGVNIIAAMP 528
Db 450 YMTSTSTPMVKYSSAMTVGSGVLSPIAFAFSRSPSLFPGILKPDIAAPGVSIILAA--507
QY 529 QNLGPTGLPYDSRRVFTVMSGTSMSCPHVSGITALIRSAYNWSPAIAKSLMTADLY 588
Db 508 ----VGDSYE-----LKSGETSMACPHVSAVALLKMWHPDWSPAIKSAIVTASVT 555
QY 589 DROGKAID--GNKPAGVPAFGAGHVPKAINPGLVYNIOVDYITVL--CTLGFTSRD 644
Db 556 DRFGMPIOAEVPRKVDAPDFGGGHIEPNKAIDPGLVYDIDPSHYTKFTFNCNLTPEADD 615
QY 645 ILAITHKWVSCNGILKRNKPFSLNPSIAVIFPKRGKTEMITRRVTVNGSPNSIYSVNVK 704
Db 616 -----CESYMEQI--YQNLPSIAVPNLKDSVT--VMRTVTVNGEAEATYHALE 661
QY 705 APEGIKVINPKRLVF-KHVQDLSYRVVMFVLKKGKRGKVASFAQGQLTWV--NSHNL 761
Db 662 APVGMTMSVEPSVITTRGSRSVTFKVTFTTTQVQGG----YTFGSLTLDGNTHS--715
QY 762 QVRSPISV 770
Db 716 --VRPIAV 722

RESULT 2

US-10-090-624-16
; Sequence 16, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472

RESULT 3

US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624

; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; BEST FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-16

Query Match 8.3%; Score 332; DB 14; Length 654;
Best Local Similarity 22.4%; Pred. No. 3.4e-22;
Matches 140; Conservative 65; Mismatches 147; Indels 274; Gaps 21;

QY 29 IVOLHNSETAKTFASKFDWHLSTFLOEAVLGVEEBEERSRLLYSYGS-----AIG 81
Db 50 IQKLNPEEISTVIV--FENH-----REKEIAVRVLELMGAKRYVVIHIIA 94
QY 82 FFAQ-----LTSEAEI--LRYSPEVAVR-----PDHVLQVOTYSYKFL 120
Db 95 IADLKVRDLVLVIGLTGKAKLSGVRFIOEDYKVTVAELEGDESAAQVMATY----149
QY 121 GLDGFNGSGWKSRRFGQGTIGVLDGTGWPSPSPDFTGMPISIPRKWGI COEGESFSS 180
Db 150 ----VWNLGYDGGTIGIITGTI-----DASHFDL-----176
QY 181 SSCNRKLGARFPFIRGHVANSPEESPNNPREVISARDSTGCTHTASTVGGSSVSMANV 240
Db 177 ---CGKVGWDFVNG-----RSY--PYDDHGHGTHVASTAAGTGAA--213
QY 241 LKNGAGVARGMAPGAHIAVYKVCWFNGCYS--SDILAAIDVAIOK---VDVLSLSLGGF 295
Db 214 ---SNKYKGMAPKALAGIKVLGADGSGSISTIIKGVWAVDNKDKYIKVINLSLSS 270
QY 296 PIPLYDDTTAIGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGATLDRRPPAVV 355
Db 271 QSSDGTALSCQVANAADAGLVVVVAAGNSGPNKYITIGSPAAASKVITVGAVDK----324
QY 356 RLANGKLYGESLYPGKGIKNAGREVVIYVGGDKGSEBCLRGSLPREIRCKWICDR 415
Db 325 -----YDV-----324
QY 416 GVNCRSEKGEAVKEAGGVAMILANTEINOEBSDIDVHLLPATLIGYTESVLLKAYNATV 475
Db 325 -----YDV-----327
QY 476 KPRARIIFGCTVIGRSRAPEVAQFSARGPSLANPSILKPDMLAPGVNIIAAMPQNLGPTG 535
Db 328 -----ITSFSSRGPTA--DGRLKPEVAVAPGNWIIAARASGTS--MG 364
QY 536 LPYDSRRVFTVMSGTSMSCPHVSGITALIRSAYNWSPAIAKSLMTADLYDROGKAI 595
Db 365 QPIND---YVTAAPGTSMATPHVAGIAULLQAPSWTPDKVXTALLETADIV-----414
QY 596 KQGNKPAGV--FAIGAGHVPKAIN 619
Db 415 ----KPDEIADIAYAGRVNAYKAIN 436

```
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRF
; ORGANISM: Thermococcus celer
US-10-090-624-12

Query Match      8.2%; Score 330; DB 14; Length 659;
Best Local Similarity 23.7%; Pred. No. 5.3e-22;
Matches 124; Conservative 57; Mismatches 119; Indels 224; Gaps 17;

QY 103 VRPHVLQVQTTYSYKFLGLDGFNGSGVSKSRP-CQGIIIGVLDGTGVWPSPSPEDDTGM 161
DB 128 IQEDYKQVDDATSVSQIGAD-----TVNLSLGYDGGVVVAIVDTGI-----DANH 174
QY 162 PSIPRWKKGICQGESFSSSSCNKRLIGARFFIRGHRVANSPEESPNMPREVISARDSTG 221
DB 175 PDL-----KG-----KVIGWYDAVNGRS-----TPYDDQG 199
QY 222 HGHTASTVGGG-SVSMANVLNGAGVARGMAPGAHIAVYKVCWFNGCY-SDILAAIDV 279
DB 200 HGTHVAGIVAGTGVNSQYI-----GVAEGAKLVGVKVLGADGSGSVTIAGVDM 250
QY 280 AIQDK-----VDVLSLGLGFPPIPLYDDTTAIGTFRAMEGIVICAGNNGPIESSVANT 335
DB 251 VQNKDKYGRVIRNLSLSSQSDGTDLSQAVNNAWDAGLVVVAAGNSGPNVTYVGSF 310
QY 336 APWVSTIGACTLDRPFPVAVRLANGKLLYGESLYPCGKIKNAGREVEIYVVTGDKGSEF 395
DB 311 AAASKVITVGAVDN----- 325
QY 396 CLRGSLPREIRKMWICDRGVNGRSEKGEAVKEAGGVAMILANTEINOEDSIDVHLLP 455
DB 326 ----- 325
QY 456 ATLIGYTESVLLKAYVNATVKPKARIIFGTVIGRGAPEVAQFSARGPSLANPSILKPD 515
DB 326 -----DNIAFSRSGPTA--DGELKPE 345
QY 516 MIAPGVNIIAAMPQNLGPT-GLYDSRRVNTVMSGTSMSCPHVSGITALIRSAYPNWSF 574
DB 346 VWAPGVDIIA--PRASGTSMTGTPIND---YYTKASGTSMATPHVSGVGAIILOAHPSWTP 400
QY 575 AAKSALMTTADLYDRGKAIKDNKPKAGVFAIGAGHVNPQKAI 618
DB 401 DRVKTALITADIV--APKEIAD-----IAYGAGRVNYYKAI 435

RESULT 4
US-10-090-624-1
; Sequence 1, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; NUMBER OF SEQ ID NOS: 33
```

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRF
; ORGANISM: Pyrococcus furiosus
US-10-090-624-1

Query Match      8.1%; Score 324.5; DB 14; Length 412;
Best Local Similarity 23.0%; Pred. No. 7.9e-22;
Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;

QY 110 QVQTTYSYKFLGLDGFNGSGVSKSRPQGTIIGVLDGTGVWPSPSPEDDTGMSIPKWK 169
DB 12 QVWATY-----VMNLGYDGGITIGIDTGI-----DASHPDL----- 44
QY 170 GICQGESFSSSSCNKRLIGARFFIRGHRVANSPEESPNMPREVISARDSTGHTGHTAST 229
DB 45 -----QGVKIVGWDFVNG-----RSY--PYDDHGHGTHVASI 74
QY 230 VGGSSVSANVLNGAGVARGMAPGAHIAVYKVCWFNGCY-SDILAAIDVAIQDK----- 284
DB 75 AAGTGAA-----SNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWEVAVDNKDKYG 127
QY 285 VDVLSLGLGFPPIPLYDDTTAIGTFRAMEGIVICAGNNGPIESSVANTAPWVSTIGA 344
DB 128 IKVINLSLSSQSDGTDLSQAVNNAWDAGLVVVAAGNSGPNKYTIGSPAAASKVITV 187
QY 345 GTLDRRPVAVRLANGKLLYGESLYPCGKIKNAGREVEIYVVTGDKGSEFCLRGSLPRE 404
DB 188 GAVDK----- 192
QY 405 EIRKQWVICDRGVNGRSEKGEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTES 464
DB 193 -----YDV----- 195
QY 465 VLLKAYVNATVKPKARIIFGTVIGRGAPEVAQFSARGPSLANPSILKPDMIAPGVNII 524
DB 196 -----ITSFSSRGPTA--DGRLEKPEVAVPAGNII 222
QY 525 AAWPQNLGPTCLPYDSRRVNTVMSGTSMSCPHVSGITALIRSAYPNWSPAKISALMTT 584
DB 223 AARASGTS-MQGPIND---YYTAAPGTSMATPHVAGIAALLLOAHPSPWDPKVTALLET 278
QY 585 ADLYDRGKAIKDNKPKAGV--FAIGAGHVNPQKAIN 619
DB 279 ADIV-----KPEIADIAYGAGRVNYYKAIN 304

RESULT 5
US-10-090-624-4
; Sequence 4, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRF
; ORGANISM: Pyrococcus furiosus
; FEATURE:
```

```

; NAME/KEY: misc_feature
; LOCATION: (428)..(428)
; OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-10-090-624-4

Query Match      8.1%; Score 324.5; DB 14; Length 522;
Best Local Similarity 23.0%; Pred. No. 1.2e-21;
Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;

QY 110 QVQTTYSYKFLGLDGFNGSVWKSRSRFGQGTIIIGVLDTGWPSPSPFDDTGMPSPRKKWK 169
Db 12 QVMATY-----VMNLGYDGGSTIIIGIITGTI-----DASHPDL----- 44
QY 170 GIQCEGESFSSSSCNRLKIGARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHAST 229
Db 45 -----QGVKIVQWDFVNG-----RSY--PYDDHGHGTHVASI 74
QY 230 VGGSSVSMANVLNGAGVARGMAPGAHIAVYKVCWFGNCGYS--SDILAAIDVAIQK----- 284
Db 75 AAGTGAA-----SNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYG 127
QY 285 VDVLISLSLGGFFPLVDYDTTIAIGTFRAMEGIGISVICAAGNNGPIESSVANTAPWYSTIGA 344
Db 128 IKVINLSLGSQSSDGTDLASQAVNAADAGLVVVVVAAGNNGPKYTTIGSPAAASKVITV 187
QY 345 GTLDRRFPVAVRLANGKLLYGESLYPGKGIKNAGREVEVIYVTGDKGSEFCILRGSLPRE 404
Db 188 GAVDK-----YDV----- 192
QY 405 EIRGKMWICDRGVNGRSEKGEAVKEAGGVAMILANTEINQEBSDIVHLLPATLIGYTES 464
Db 193 ----- 195
QY 465 VLLKAYNATVKPKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNII 524
Db 196 -----ITFSRSGPTA--DGRLLPEVYVAPGNWII 222
QY 525 AAMPQNLGPTGLPYDSRRVNTVMSGTSMCPHVSGITALIRSAVFNWSPAALKALMTTADLY 584
Db 223 AARASGTS--MGQPLND---YTTAAPTGMTATPHVAGVALLQAHQHPSTWPKVKYKIALIET 278
QY 585 ADLYDRQCKAIDCKNKPAGV--FAIGAGHVNPQKAIN 619
Db 279 ADIV-----KPEDEADIAYGAGRVNAYKAIN 304

RESULT 6
US-10-156-761-13251
; Sequence 13251, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13251
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13251

Query Match      7.9%; Score 318.5; DB 15; Length 1208;
Best Local Similarity 24.5%; Pred. No. 1.9e-20;
Matches 129; Conservative 45; Mismatches 113; Indels 239; Gaps 19;

QY 126 GNSGVWSKSRFGQGTIIIGVLDTGWPSPSPFDDTGMPSPRKKWKIGICQEGESFSSSSCNR 185
Db 198 GTRAANDAGUTGQVTVAVUDTG-----DTHPDL-----AGVRSRKS--- 237
QY 186 KLIGARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVSMANVLNGA 245
Db 238 -----FIDGEEVA-----DRNGHGHGTHVSTVGGSGAA---SD 266
QY 246 GVARGMAPGAHIAVYKVCWFGNCGYS--SDILAAIDVAIQD--KVDVLSLSLGGF-----P 296
Db 267 GTERGVAPGATLAVGKVLSDQAGAGSESOIAGMEWAARDVRAIRIVMSLSGSTEASDGTDP 326
QY 297 IPLYDDTIAIGTFRAMEGIGISVICAAGNNGPIESSVANTAPWYSTIGAGTLDRRFPVAVR 356
Db 327 MAEAVDTL-----SEETGALFVVAAGNTG-----AP--SSTGS----- 357
QY 357 LANGKLLYGESLYPGKGIKNAGREVEVIYVTGDKGSEFCILRGSLPREIRGRMWICDRG 416
Db 358 -----PGAA----- 361
QY 417 VNGRSEKGEAVKEAGGVAMILANTEINQEBSDIVHLLPATLIGYTESVLLKAYNATVK 476
Db 362 -----DS-----ALTVGAVDS----- 372
QY 477 PKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGL 536
Db 373 -----SDRAAYFTSAGFRHGD--NALKPDLAAPGVDIRAARSOLAPGTGY 415
QY 537 PYDSRRVNTVMSGTSMCPHVSGITALIRSAVFNWSPAALKALMTTADLYDRQCKAIAK 596
Db 416 -----YTSMTSTWATPHVAGVALLAEQHPDWTGARKLDMSTSQLD----- 460
QY 597 DGNKPAGVFAIGAGHVNPQKAINPGLVNIQPVYIITLCTLGTR 642
Db 461 -----ASYVQLGAGRVSPDAGVART-----ATGSADLGFHR 493

RESULT 7
US-10-156-761-10856
; Sequence 10856, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10856
; LENGTH: 1139
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10856

Query Match      7.6%; Score 307; DB 15; Length 1139;
Best Local Similarity 23.2%; Pred. No. 2.1e-19;
Matches 153; Conservative 58; Mismatches 192; Indels 256; Gaps 22;

QY 126 GNSGVWSKSRFGQGTIIIGVLDTGWPSPSPFDDTGMPSPRKKWKIGICQEGESFSSSSCNR 185
```

Db 241 CAPTAWAGYDGKGYKAVLDTGVDATPDLKQ-----VAESKNFSAAA--- 285
Qy 156 KLIGARPPIRHVRANSPEESPNNPREVISARDSTGHGTHASTVGGSSVSMANVLNGA 245
Db 286 -----DAADFHGTHVASTAAGTG-----AKSN 309
Qy 246 GVARGMAPGAHIAVYKVCWFNGC-YSSDILAAIDVAIQDKVDVLSLSLGGFPPIPLYDDTI 304
Db 310 KKYGVAPGATILNKVLDLDTGSDGSLAGMEWAAPQAGADVNLVSLGGDTPEIDPLE 369
Qy 305 AIGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGACTLDRRPPAVVRLANGKLLY 364
Db 370 AEVNLKSEKGIILFAIAGNEGE-----F 393
Qy 365 GESLYPGKGIKNAGREVEVIYVGGDKSEFCLESLPREIRKMWICDRGVNGRSEK 424
Db 394 GE----- 395
Qy 425 EAVKEAGGVAMILANTEINBEDSIDVHLLPATLIGYTESVLLKAYVNVATVKPKARIIFG 484
Db 396 QTIGSPGAADALTGAVNDS----- 417
Qy 485 GTVIGRSAPRVAQFSARGPSLANPSILKPDMIAPGVNIIAA-----WPNLGL--PTGL 536
Db 418 -----KLASFSSRGPL--DGAIKPDVTAPGVDITAAAPGVSVIDQEVQKPDG- 464
Qy 537 PYDSRRVNTVSGTSMSCPHVSGITALLIRSAYNWSPAAIKSALMTTADLYDROGKAIK 596
Db 465 -----YLTISGTSNATPHVAGAAAILKQHPNWSFAELKALTGSA-----KGGKYTP 512
Qy 597 DGNKPAGVFAIGAHVNPQKAINPGLVNIQPDY-----ITVLCITLGF 640
Db 513 -----FOGSGRIADVKAIKOSVTANPNSVSGTQOHPHTDDKPVTOQLTVR--NLG- 582
Qy 641 TRSDI---LAITHKNVSCGILRKNPGFSLNYPGSIIVIFKRGKTTMITRRVTNVGSPNS 697
Db 563 -TSDVTLNASTATNP--KGVAASGPFKLGATKVTVP--AGKASVDFTVNTKLGTTDG 618
Qy 698 IYVNVKAPGKIVNPKRLVFKHVDQTLRYVNVFVKKNRGKGKVASFAQGLTWN 756
Db 619 AYSAVVTATGGQT--VRTAAAVQREVE---SYDV--TLKHIDRDKPANYSTDLTGVS 671

RESULT 8

US-10-112-488-39
; Sequence 39, Application US/10112488
; Publication No. US20030082746A1
; GENERAL INFORMATION:
; APPLICANT: KIKUCHI, Yoshiaki
; APPLICANT: DATE, Masayo
; APPLICANT: UMEZAWA, Yukiko
; APPLICANT: YOKOYAMA, Keiichi
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
; FILE REFERENCE: 219286USOCNT
; CURRENT APPLICATION NUMBER: US/10/112,488
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: PCN/JPO0/06780
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: JP2000-280098
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: JP11-280098
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Streptomyces alboborisolus
US-10-112-488-39

Query Match

7.0%; Score 282; DB 15; Length 1079;

Best Local Similarity 20.8%; Pred. No. 4.8e-17;
Matches 158; Conservative 71; Mismatches 257; Indels 272; Gaps 22;
Qy 36 SETAKTFAS-KPDHLSFLQEAVALGVEEEEPSSRLLYSYCSAIEGFAAQ----- 85
Db 67 ADARILVAGKLDQRLFDVTELNKAATTAHKGKLVIVGVYGAKAADVRDAGTVRR 126
Qy 86 -LSESEAEILRYSPEVAVRDPHVLQVQTTYS-YKFLGLDG-----FNSGVW 131
Db 127 TLTSLNADAVQTPQAGAELEWAVTDGRTASGVARVWLDGVRKASLDTSVGQIGTPKAW 186
Qy 132 SKSRFGQGTIIQVLTGTVWPSFSPDDTGMPISIPRKWGIQCEGSEFSSSSCNRLKIGAR 191
Db 187 EAGYDGKGYKIAVLDTG-----DATHPDL-----KGQVTASKNFTSAP----- 225
Qy 192 FIIRHVRANSPEESPNNPREVISARDSTGHGTHASTVGGSSVSMANVLNGAGVARGM 251
Db 226 -----TTGDVVGHGTHVASTAAGTG-----AQSKTYTKGV 255
Qy 252 AFGAHIAVYKVCWFNGC-YSSDILAAIDVAIQDKVDVLSLSLGGFPPIPLYDDTIAGTFR 310
Db 256 AFGAKILNGKVLDDAGFGDDSGILAGMEWAAGADIWNMSLGGWDTTETDPLEAAVDKL 315
Qy 311 AMERGISVICAAGNNGPIESSVANTAPWVSTIGACTLDRRPPAVVRLANGKLLYGESLYP 370
Db 316 SAEKGIILFAIAGNEGPOSIGSPGSADSALTGVA----- 349
Qy 371 GKGIKNAGREVEVIYVGGDKSEFCLESLPREIRKMWICDRGVNGRSEKGEAVKEA 430
Db 350 -----VDDKDK----- 355
Qy 431 GGVAMILANTEINBEDSIDVHLLPATLIGYTESVLLKAYVNVATVKPKARIIFGTVIGR 490
Db 356 ----- 355
Qy 491 SRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGLPYDSRRVNTVMSG 550
Db 356 -----LADFSSTGPRLGDAV-KPDLTAPGVDITAAASAG-NDIAKEVGEKPAGYMTISG 408
Qy 551 TSMSCPHVSGITALLIRSAYNWSPAAIKSALMTTADLYDROGKAIKDGNKPKAGVFAIGAG 610
Db 409 TSNATPHVAGAAAILKQHPENKAYELKALTAST-----KDGKYTP--FEQSG 456
Qy 611 HYNPKAINPGLVNIQPDYITVLCITLGF-----TRSDILAIATHKNVSCGILRK 661
Db 457 RVQVDKAITQTVI--AEPV-----SLSFGVQOHPHADDDKEVTKKLYRNLTGDEDTLK 507
Qy 662 -----NPGFSLNYPGSIIVIFKRGKTTMITRRVTNVGSPNSIYVNVKAPGEG 708
Db 508 LTSTATGPKGAAPAGFFTLGASTLTVPANGTASVDVTADTLGAVDGTYSAYVVA-TG 566
Qy 709 IKVINPKELVFKHVDQTLRYVNVFVKKNRGKGKVAS 746
Db 567 AQQSVRTAAAVEREVE---SYNV--TLKVLDRSGKATA 599

RESULT 9

US-09-813-408-7
; Sequence 7, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Marrs, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 382
; TYPE: PRT

```
; ORGANISM: Bacillus sp.
US-09-813-408-7

Query Match      6.9%; Score 276.5; DB 11; Length 382;
Best Local Similarity 20.8%; Pred. No. 2.7e-17;
Matches 127; Conservative 62; Mismatches 177; Indels 245; Gaps 21;

QY 8 LCIIIFLLFCSSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSEFLQEAVLGVVEEBEE 67
DB 14 LLLSILTSATSVSAEEQKKVLIGFE-NQLQVTEFVSSDKGOS---EMSLFAEVNDESI 69
QY 68 SSRLLYSYGAIEGFAAQLTESEAEILRYSPVAVRPHDLVQLQVQTTYSYKFLGLDGF 127
DB 70 EMELLYEF-EDIPVVSVELSPEDVKLEKXPSITYIBED--IEVTITNQVTPWGITRVOA 126
QY 128 SGVSKRPFQGTIIIGVLDTCWMPESPFDGTGMPSPRKKWKGICQEGESFSSSSCNRL 197
DB 127 PTAWTRGTGTGVRVAVLDGTGI-STHPDLNIRG-----GVSF----- 162
QY 188 IGARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVSMANVLNGAGV 247
DB 163 -----VPGEP-----SYQDNGHGTHVAGTIA-----ALNNSIGV 192
QY 248 ARGMAPGAHIAVYKVMFNGCYS-SDILAAIDVAIQKVDVLSLGGFFPIPLYDDTIAI 306
DB 193 V-GVAPNAELVAVKVLGANGSGSVSSIAQGLQWTAQNNIHVANLSLGS---PVGSQTL 248
QY 307 GTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTILDRFPVAVRLANGKLLY 366
DB 249 AVNQATNAGVLVAATGNN-----GSGTVS----- 273
QY 367 SLYPCKGIGKNAGREVEVIYVTGGDKGSEFCLRGSLPREIRGKMWICDRGVNGRSEK 456
DB 274 ---YPAR----- 277
QY 427 VKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYVNATVKPKARIIFG 486
DB 278 -----YANA----- 287
QY 487 VIGSRAPVAVQFARGPSLANPSILKPDMTAPGVNIIAAMPQNLGTLGTPYDSRRVNF 546
DB 288 DQNNR-----ASFQYGTGL-----NIVAPGVIGQSTVPGN-----RYA 322
QY 547 VMSGTSMSCPVSGTALIRSAYPNWSPAIKSALMTTADLYDRQKAIKDGKPKAGVFA 606
DB 323 SLSGTSMATPHVAGVAAVVKQNPWSWNTQIRQHLTSTA-----TSLGNSNQ----- 369
QY 607 IGAGHVNPQKA 617
DB 370 FGSGLVNAEAA 380

RESULT 10
US-10-146-905A-8
; Sequence 8, Application US/10146905A
; Publication No. US20030215906A1
; GENERAL INFORMATION:
; APPLICANT: Boon Leong Lim
; TITLE OF INVENTION: Recombinant Bacillus Proteases and Uses Thereof
; FILE REFERENCE: 9661-035-999
; CURRENT APPLICATION NUMBER: US/10/146,905A
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-10-146-905A-8

Query Match      6.8%; Score 271.5; DB 12; Length 382;
Best Local Similarity 20.5%; Pred. No. 8.1e-17;
Matches 126; Conservative 63; Mismatches 177; Indels 249; Gaps 17;

QY 6 FFLCIIFLLFCSSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSEFLQEAVLGVVEEBEE 65
DB 12 FALALIFTMFGSTSS-----AOAGKSGNGEKYIVGFQKTMSTMSAAKKXDVISEK 63
QY 66 EPSSRLYSYGAIEGFAAQLTESEAEILRYSPVAVRPHDLVQLQVQTTYSYKFLGLD 124
DB 64 GKVQKQKPY---VDAASATLNKAVKELKDPVAVVEEDHVAH---AYAQSVPYGVUSQ 117
QY 125 FGNVGWMSKRFQGTIIIGVLDTCWMPESPFDGTGMPSPRKKWKGICQEGESFSSSSCN 184
DB 118 IKAPALHSGQYTGNSVKNVAVIDSDSHPLDKVAGGASM-----ALNNS 157
QY 185 RKLIGARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVSMANVLNG 244
DB 158 -----VPSETNPFQDNNHGHVAGTVA-----ALNNS 185
QY 245 AQVARGMAPGAHIAVYKVMFNGC-YSSDILAAIDVAIQKVDVLSLGGFFPIPLYDDT 303
DB 186 IGVL-GVAPASLAVKVLGADSGQYSWIIINGEIAIANMMDVINMSLGG---PSGSAA 241
QY 304 IAIGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTILDRFPVAVRLANGKLL 363
DB 242 LKAAYDKAVAGVVVVAAGNEGTSGSS-----STVG----- 273
QY 364 YGESLYPGKIGKNAGREVEVIYVTGGDKGSEFCLRGSLPREIRGKMWICDRGVNGRSEK 423
DB 274 -----YFGK----- 277
QY 424 GEAVKEAGGVAMILANTEINOEDSIDVHLLPATL-IGYTESVLLKAYVNATVKPKARI 482
DB 278 -----YPSVIAVGAVDS----- 289
QY 483 FGGTVIGSRAPVAVQFARGPSLANPSILKPDMTAPGVNIIAAMPQNLGTLGTPYDSRR 542
DB 290 -----SNQRASFSSVGPGL-----DVNAPGVSIQSTLPGN----- 319
QY 543 VNFTVMSGTSMSCPVSGTALIRSAYPNWSPAIKSALMTTADLYDRQKAIKDGKPKA 602
DB 320 -KYGAYNGTSMASPHVAGAAAILSKHENWINTQVRSSLENT-----TKLG 365
QY 603 GVFAIGAGHVNPQKA 617
DB 366 DSFFYVGKGLINQAA 380

RESULT 11
US-10-090-624-31
; Sequence 31, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-090-624-31

Query Match      6.8%; Score 271.5; DB 14; Length 382;
```

Best Local Similarity 20.5%; Pred. No. 8.1e-17;
Matches 126; Conservative 63; Mismatches 177; Indels 249; Gaps 17;

6 PFLCIIIFLLFCSSSEILQKOTYIVQLHPNSETAKTFASKFDWHLSPLOBAVLGVVEEEE 65
12 PALALIFTMARGSTSS-----AQAAGKNGEKKYIVGFKQTMSTWSAAKKKDVISEK 63
66 EPSRLLYSYGSAIEGFAAQTESAEILRYSPVAVRDPHVLQVOTTYSYKF-LGLDG 124
64 GKGKQKQFYK---VDAASATLNEKAVKELKDPSPVAYVEEDHVAH---AYAQSPVYGVQS 117
125 FGNSSGWSKSPFGQGTIGVLDGTWVPESPSPDDTGMPSIPRKWKGICQEGESFSSSSCN 184
118 IKAPALHSQGYTGNNKVAVIDSGIDSHHPLKVVAGASM----- 157
185 RKLIGARFFIRGHRVANSPESSPNPREYISARDSTGHGTHTASTVGGSSVSMANVLNG 244
158 -----VPSETNPFQDNNSHGTHVAGTVA-----ALNNS 185
245 AGVARGMAPGAHIAVYKVCWFGNC-YSSDILAAIDVAIQDKVDVLSLGLGFPPIPLYDDT 303
186 IGV-L-GVAPSASLYAVKVLGADGSGQYSWIINGIEMAIANNMVDVINMSLGG---PSGSAA 241
304 IAGTFRAMERGISVICAGNNGPIESSVANTAPWVSTIGAGTLDLDRPPAVVRLANGKLL 363
242 LKAAVDKAVASGVVVAAGNEGTSGSS-----STVG----- 273
364 YGESLYPGKGIKNAGREVEVIYVGGDKGSBFLRGLSPREEIRGKMWICDRGVNCRSEK 423
274 -----YPGK----- 277
424 GEAVKEAGGVAMILANTEINOEDSIDVHLLPATL-IGYTESVLLKAYVNAVTKPKARI 482
278 -----YPSVIANGVDS----- 289
483 FGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAFGVNIIAAMPONLQPTGLPYDSRR 542
290 -----SNQASPSVSGPEL-----DVMAFGVSIQSTLPGN----- 319
543 VNFVMSGTSMSCHVSGITILIRSAYPNWSPPAAIKSALMTADLYDRQKAIKDGKPA 602
320 -KYGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENT-----TKLG 365
603 GVFAIGAGHVNPKA 617
366 DSFFYKGLINVOAA 380

RESULT 12
US-10-104-693-2
; Sequence 2, Application US/10104693
; Publication No. US20030118605A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; APPLICANT: Ganshaw, Grant C.
; APPLICANT: Harding, Fiona A.
; APPLICANT: Larenas, Edmund A.
; APPLICANT: Poulos, Ayrockazan J.
; APPLICANT: Sikorski, Elizabeth E.
; APPLICANT: Russell, Elliott P.
; TITLE OF INVENTION: Proteins Producing an Altered Immunogenic
; FILE OF INVENTION: Response and Methods of Making and Using the Same
; FILE REFERENCE: GC683-2
; CURRENT APPLICATION NUMBER: US/10/104,693
; CURRENT FILING DATE: 2002-08-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus amyloiquefaciens
US-10-104-693-2

Query Match 6.8%; Score 271.5; DB 15; Length 382;
Best Local Similarity 20.5%; Pred. No. 8.1e-17;
Matches 126; Conservative 63; Mismatches 177; Indels 249; Gaps 17;

6 PFLCIIIFLLFCSSSEILQKOTYIVQLHPNSETAKTFASKFDWHLSPLOBAVLGVVEEEE 65
12 PALALIFTMARGSTSS-----AQAAGKNGEKKYIVGFKQTMSTWSAAKKKDVISEK 63
66 EPSRLLYSYGSAIEGFAAQTESAEILRYSPVAVRDPHVLQVOTTYSYKF-LGLDG 124
64 GKGKQKQFYK---VDAASATLNEKAVKELKDPSPVAYVEEDHVAH---AYAQSPVYGVQS 117
125 FGNSSGWSKSPFGQGTIGVLDGTWVPESPSPDDTGMPSIPRKWKGICQEGESFSSSSCN 184
118 IKAPALHSQGYTGNNKVAVIDSGIDSHHPLKVVAGASM----- 157
185 RKLIGARFFIRGHRVANSPESSPNPREYISARDSTGHGTHTASTVGGSSVSMANVLNG 244
158 -----VPSETNPFQDNNSHGTHVAGTVA-----ALNNS 185
245 AGVARGMAPGAHIAVYKVCWFGNC-YSSDILAAIDVAIQDKVDVLSLGLGFPPIPLYDDT 303
186 IGV-L-GVAPSASLYAVKVLGADGSGQYSWIINGIEMAIANNMVDVINMSLGG---PSGSAA 241
304 IAGTFRAMERGISVICAGNNGPIESSVANTAPWVSTIGAGTLDLDRPPAVVRLANGKLL 363
242 LKAAVDKAVASGVVVAAGNEGTSGSS-----STVG----- 273
364 YGESLYPGKGIKNAGREVEVIYVGGDKGSBFLRGLSPREEIRGKMWICDRGVNCRSEK 423
274 -----YPGK----- 277
424 GEAVKEAGGVAMILANTEINOEDSIDVHLLPATL-IGYTESVLLKAYVNAVTKPKARI 482
278 -----YPSVIANGVDS----- 289
483 FGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAFGVNIIAAMPONLQPTGLPYDSRR 542
290 -----SNQASPSVSGPEL-----DVMAFGVSIQSTLPGN----- 319
543 VNFVMSGTSMSCHVSGITILIRSAYPNWSPPAAIKSALMTADLYDRQKAIKDGKPA 602
320 -KYGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENT-----TKLG 365
603 GVFAIGAGHVNPKA 617
366 DSFFYKGLINVOAA 380

RESULT 13
US-09-870-122-3
; Sequence 3, Application US/09870122
; Patent No. US20020142009A1
; GENERAL INFORMATION:
; APPLICANT: Regents of the University of Minnesota et al.
; TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
; FILE REFERENCE: 600.450W01
; CURRENT APPLICATION NUMBER: US/09/870,122
; CURRENT FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: US 09/206,898
; PRIOR FILING DATE: 1998-12-07
; PRIOR APPLICATION NUMBER: US 08/589,756
; PRIOR FILING DATE: 1996-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1150
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-870-122-3

Query Match 6.7%; Score 269.5; DB 10; Length 1150;
Best Local Similarity 24.2%; Pred. No. 8.3e-16;

```
Matches 128; Conservative 92; Mismatches 201; Indels 107; Gaps 25;
QY 133 KSPFGQGTIIICVLDGTGWPEPSFDDTGMPS-STPRKWKICQEGESFSSSSCNKRKLIGAR 191
Db 117 KAGKAGTVAVIDAG-----FDKNHEAWRLTDKTARYOSKEDLEKA---KKEHGIT 166
QY 192 FFIRGHRVANSPEESPNMPEYISARDSTGHGTHHTASTVGGSSVSMANVLGNGAGVARGM 251
Db 167 Y---GEWNDKVAYHYDYSKDGTAVDQOE-HGTHVSGILSGNAPSTKEPYR-----LEGA 218
QY 252 APGAHIAVYKVCWFNGC--YSSDILAAIDVAIQDKVDVLSLGGFPPI---LYDDTIAI 306
Db 219 MPEAQILLMRVEIVNGLDVARNYAQAIRDAINLGAKVINMSFGNAALAYANLPDETCKA 278
QY 307 GTFRAMEGHSVCAAGNG-----PI-----ESSVANT---APWVSTIGAGTLDREF 351
Db 279 FDY-AKSGVSIIVTSAGNDSFGKTRPLADHPDYGVWGTTPAADSTLTIVASYSPOKOL 337
QY 352 PAVURLANG-----KLLYGESLYPGKGIKNAGREVEVIYVGTGPKGSEFCLRGSLPRE 404
Db 338 TETVRVTACQOKKEMPEVLTNRPENKAYDYA-----YANRGTKEDDF-----K 382
QY 405 EIRGKMWICDRGVNGRSEKGEAVKEAGGVAMILANTEINOQE-----DSIDVHLLPATLIG 460
Db 383 DVKGKIALIERGDDIDFKDKIAKAKAGAGVGLIYD---NQDKGFPTELPNVQMPAARTS 439
QY 461 YTESVLKAYVNAVTKPKARIIFGGT--VIGRSRAPEVAOFSARGPSLANPSILKXPDMA 518
Db 440 RKDGLLEKD-----NPQKITFNATPKVLPTASGTKLSRFSSWG--LTADGNIKEDIAA 491
QY 519 PGVNIIAWPNQLGPTGLPYDSRRVNTVMSGTSMSCPHVSIGITALIR-----SAYPNWSP 574
Db 492 PGQDILSSVANN-----KYAKLSGTSMSAPLVAGIWMGLLQKQYETQYDPDMTP 538
QY 575 A----AIKSAALMTTAD-LYDRQKKAIKDGNKPAGVFAIGAGHVNPOKA 617
Db 539 SERLDLAKVLMSSATLYDEDEKAYESPQQ-----GAGAVDAKKA 580

RESULT 14
US-10-146-905A-11
; Sequence 11, Application US/10146905A
; Publication No. US20030215906A1
; GENERAL INFORMATION:
; APPLICANT: Boon Leong Lim
; TITLE OF INVENTION: Recombinant Bacillus Proteases and Uses Thereof
; FILE REFERENCE: 9661-035-999
; CURRENT APPLICATION NUMBER: US/10/146,905A
; CURRENT FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion protein between the signal peptide from B. subtilis
; and BPN' protease from B. amyloliquefaciens
US-10-146-905A-11

Query Match 6.6%; Score 266.5; DB 12; Length 382;
Best Local Similarity 20.5%; Pred. No. 2.4e-16;
Matches 127; Conservative 65; Mismatches 168; Indels 261; Gaps 20;
QY 6 PFLCIIF-LFCCSSSSSEILQ-----KQYIIVQLHPNSETAKTPASKFDWHLSPFQEAVALG 59
Db 12 FALTLLIFTMAFNSQAQAAGKNGEKYIVGFKQTMSTMSA-AKKKD-----V 59
QY 60 VEEBEEPSRLLYSYGSAIEGFAAQLTESEAEILRYSPEVAVRDPHVLQVQVTTYKYKF 119
Db 60 ISEKGGKVQKQFKY-----VDAASATLNEKAVKELKXDPVAYVEEDHVAH---AYAQSV 111
QY 120 -LGLDGFNGSWGSKSRFQGTIIIGVLDGTGWPEPSFDDTGMPSIPRKWKICQEGESF 178
```

```
Db 112 PYGYSQIKAPALHSQGYTGSNVKVAVIDSGIDSSHDPDLKVAGGASM----- 157
QY 179 SSSSCNKRKLIGARFFIRGHRVANSPEESPNMPEYISARDSTGHGTHHTASTVGGSSVMA 238
Db 158 -----VPSETNPFQDNNSHGTHVACTVA----- 180
QY 239 NVLNGAGVARGVAPGAHIAVYKVCWFNGC--YSSDILAAIDVAIQDKVDVLSLGGFPPI 297
Db 181 -ALANSIGVL-GVAPSASLYAVKVLGADGSGQYSWIINGIEWAIANNMDVINMSLGG--- 235
QY 298 PLYDDTIAIGTFRAMEGHSVCAAGNGPIESVANTAPWVSTIGAGTLDREFPAVURL 357
Db 236 PSGSAALKAADVAKAVAGSVVVAAGNEGTSGSS-----STVG----- 273
QY 358 ANGKLLYGESLYPGKGIKNAGREVEVIYVGTGDKGSEFCLRGSLPREIRGKMWICDRGV 417
Db 274 -----YPOK----- 277
QY 418 NGRSEKGEAVKEAGGVAMILANTEINOQEEDSIDVHLLPATL-IGYTESVLLKAYVNAVTK 476
Db 278 -----YPSVIAVGAVD----- 289
QY 477 PKARIIFGGTVIGRSRAPEVAOFSARGPSLANPSILKXPDMAIPGVNIIAAMPQNLGPTGL 536
Db 290 -----SNQRASFSSVGPPEL-----DYMAGVSIQSTLPGN----- 319
QY 537 PYDSRRVNTVMSGTSMSCPHVSIGITALIRSAVFNWSPAAIKSALMTTADLYDRQKKAIK 596
Db 320 -----KYGAYNTGTSMASPHVAGAAALILSKHPNWTNTQVRSLNLT----- 361
QY 597 DGNKFPAGVFAIGAGHVNPOKA 617
Db 362 --TKLGDSEFYVGKLIINVQAA 380

RESULT 15
US-09-813-408-20
; Sequence 20, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Marrs, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-09-813-408-20

Query Match 6.6%; Score 264.5; DB 11; Length 382;
Best Local Similarity 20.2%; Pred. No. 3.8e-16;
Matches 126; Conservative 68; Mismatches 168; Indels 263; Gaps 18;
QY 6 PFLCIIF-LFCCSSSSSEILQKQYIIVQLHPNSETAKTPASKFDWHLSPFQEAVALGV EEE 65
Db 6 FALALIFTMAFGSTSS-----AQAAKNGEKYIVGFKQTMSTMSA-AKKKDVISEK 57
QY 66 EPSEBLLYSYGSAIEGFAAQLTESEAEILRYSPEVAVRDPHVLQVQVTTYKYKGLDGF 125
Db 58 GKGVQKQFKY---VDAASATLNEKAVKELKXDPVAYVEEDHV-----AHAYAQSVPY 107
QY 126 GNSGVMSKSRFQGG-----TIIGVLDGTGWPEPSFDDTGMPSIPRKWKICQEGESF 178
Db 108 GVSQIKAPALHSQGYTGSNVKVAVIDSGIDSSHDPDLKVAGGASM----- 153
QY 179 SSSSCNKRKLIGARFFIRGHRVANSPEESPNMPEYISARDSTGHGTHHTASTVGGSSVMA 238
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 00:24:01 ; Search time 376 Seconds
(without alignments)
5564.001 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEKPFLLCIPLFLFCSSSS.....NSHNLQVRSPISVTLKTN 775

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09806767/runat_02022004_085957.17887/app_query.fasta_1.967
-DB=N Geneseq 19Jun03 -OPMT=fastap -SUFFIX=xng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09806767 @CIGN 1.1.0 @runat_02022004_085957.17887 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03:**

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:**
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:**
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:**
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:**
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:**
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:**
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:**
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:**
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:**
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:**
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:**
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:**
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:**
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:**
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:**
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:**
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:**
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:**
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:**
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:**
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:**
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:**
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:**
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:**
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4018	100.0	2328	21	AAZ93808 SDD1 sequence enco
2	4009	99.8	2328	21	AAZ93809 SDD1 sequence enco
3	2786.5	69.4	3140	21	AAZ93814 Subtilase (st p2sc
4	2727.5	67.9	2492	21	AAZ93813 Subtilase (Plgsep)
5	2531.5	63.0	3865	21	AAZ93812 Subtilase (Plgsep)
6	1742	43.4	2265	21	AAZ93817 Arabidopsis thalia
7	1740	42.3	2265	21	AAZ93816 Arabidopsis thalia
8	1700	42.3	2637	21	AAZ93815 Arabidopsis thalia
9	1684	41.9	2638	21	AAZ93814 Arabidopsis thalia
10	1642.5	40.9	2434	21	AAZ93813 Arabidopsis thalia
11	1589	39.5	2295	24	ABZ13381 Arabidopsis thalia
12	1589	39.5	2489	21	AAZ93812 Arabidopsis thalia
13	1214	30.2	2402	21	AAZ93811 Arabidopsis thalia
14	1172.5	29.2	2552	15	AAQ73756 Gene encoding melo
15	1095.5	27.3	2448	20	AAZ93905 cDNA encoding a SC
16	1074	26.7	2958	24	ABZ12533 Arabidopsis thalia
17	1070	26.6	2310	25	ABZ42107 Arabidopsis thalia
18	1056.5	26.3	1653	21	AAZ93808 Arabidopsis thalia
19	909	22.6	513445	22	AAI61373 Soybean 318013 reg
20	901	22.4	2451	24	ABZ12259 Arabidopsis thalia
21	806	20.1	7235	20	AAZ93910 Genomic DNA encodi
22	623.5	15.5	740	24	ABQ5482 Arabidopsis thalia
23	448.5	11.2	975	21	AAZ93808 Arabidopsis thalia
24	430	10.7	2532	13	AAQ91334 Encodes RP-III res
25	388	9.7	436	25	ABX61910 Arabidopsis thalia
26	387	9.6	1674	24	ABK78984 Bacillus clausii G
27	345	8.6	1977	18	AAZ93869 Protease coding se
28	332	8.3	1962	18	AAZ93869 Pyrococcus furiosu
29	332	8.3	1962	20	AAZ93869 Hyperthermostable
30	330	8.2	1977	18	AAZ93869 Thermococcus prote
31	330	8.2	1977	20	AAZ93869 WO9856926 Seq ID 1
32	324.5	8.1	1236	20	AAZ93869 Hyperthermostable
33	324.5	8.1	1566	18	AAZ93869 Pyrococcus furiosu
34	322.5	8.0	7156	19	AAV15586 Lactobacillus bulg
35	297	7.4	2539	18	AAZ93869 Streptomyces virid
36	297	7.4	2809	18	AAZ93869 DhpA-mel chimeric
37	296.5	7.4	4740	25	ABZ58957 Group B Streptococ
38	291.5	7.3	2121	23	ABZ58957 T. yonsei sublini
39	291.5	7.3	4941	24	ABN69192 Streptococcus poly
40	289	7.2	528	25	ABX57523 Arabidopsis thalia
41	288.5	7.2	2655	24	ABN67963 Streptococcus poly
42	288.5	7.2	2655	24	ABN70427 Streptococcus poly
43	288.5	7.2	2155561	24	ABN71527 Streptococcus poly
44	287	7.1	343	25	ABX21785 Human GDP-mannose
45	285	7.1	4650	24	ABN71526 Streptococcus poly

ALIGNMENTS

RESULT 1

AAZ93808
ID: AAZ93808 standard; DNA; 2328 BP.

XX AC AAZ93808;

XX DT 16-AUG-2000 (first entry)

XX SDD1 sequence encoding a subtilisin like serine protease.

XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
XX stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
XX crop protection; feed; foodstuffs; ss.

XX Arabidopsis thaliana.

XX Key

Location/Qualifiers

FH


```

Db 1441 ATAAATTTTGGTGGTACGGTGATTGGAGAGTACAGACACCGAGGTGGCTCAGTTTCA 1500
Qy 501 AlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGly 520
Db 1501 GCTCGAGGACCGAGTTTACCAATCTTCGATACATAAACCAGGATATGATTCTCGGGA 1560
Qy 521 ValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSer 540
Db 1561 GTCAATATCATCTCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTTATGATTC 1620
Qy 541 ArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGly 560
Db 1621 AGAAGAGTTAACTTCACTGTAATGTCCAGAACTCAATGTCTGTCCACATGTTAGCGGA 1680
Qy 561 IleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAla 580
Db 1681 ATCAGTGTCTTATCCGCTCTGATACCCGAACTGCTCCAGCTGCAATCAAAATCCGGA 1740
Qy 581 LeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLeAlaIleLysAspGlyAsnLys 600
Db 1741 TTGATGACACACGCGATTGTAGCATGTCAAGGGAAGCGATAAAGGATGTAACAAA 1800
Qy 601 ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 1801 CCAGCCGGTGTGTTCGATTTGGAGCAGGCGCATGTCAATCCGCAAAAGCGATAAACCG 1860
Qy 621 GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPhe 640
Db 1861 GGAATTGGTTTACAAACATTCAACAGTGGATTACATAAATCTTACCTCTGGATTTC 1920
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 1921 ACAAGATCATGATTTTATGATCATCACTATAGAACGTGAGCTGCAATGGAATATGCGG 1980
Qy 661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 1981 AAAACCCGGTGTAGTCTCAATTACCGTGCATAGCCGTGATTTTCAACGTCGCAAG 2040
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSer 700
Db 2041 ACTACGAGATGATCACAGCGGTGTCACTAAAGTTGGGAGTCTAACTCGATATACTCA 2100
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db 2101 GTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTATTGTCAATCCTAAGAGACTTGTGTT 2160
Qy 721 LysHisValAspGlnThrLeuSerTyrArgValTrpPheValLeuLysLysLysAsnArg 740
Db 2161 AAACACGTGGATCAGACGCTGAGCTATAGATATGGTTGTATGAGAGAGAAACACAGA 2220
Qy 741 GlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeu 760
Db 2221 GGAGGGAAGTGGCTAGCTTTGCACAAAGGCACTTGACTTGGGTCAACTCTCATATATCG 2280
Qy 761 MetGlnArgValArgSerProIleSerValThrLeuLysThrAsn 775
Db 2281 ATCAGCGAGTTAGAAGTCCAAATCTGTAAACCTTGAAGACTAAC 2325

```

RESULT 2

AAZ93809

ID AAZ93809 standard; DNA; 2328 BP.

AC

AAZ93809;

DT 16-AUG-2000 (first entry)

XX

DE SDD1 sequence encoding a subtilisin like serine protease.

XX

KW SDD1; serine protease; subtilisin; transgenic plants; dry weight;

KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;

KW crop protection; feed; foodstuffs; ss.

XX

```

OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 1..1476
FT /*tag= a
FT /product= SDD1 subtilisin-like serine protease.
XX
FN WO200022144-A2.
XX
PD 20-APR-2000.
XX
PF 12-OCT-1999; 99WO-EP07633.
XX
PR 12-OCT-1998; 98EP-0119244.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
PI Berger D, Altmann T;
XX
DR WPI; 2000-317995/27.
XX P-PSDB; AAY83301.
XX
PT Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
PT lower water consumption and enhanced diseased resistance
XX
PS Claim 1; Page 66-68; 101pp; English.
XX
CC Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO2 uptake into and H2O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This sequence is
CC identical to the one given in GENESEQ record AAZ93808 with the
CC exception of a single point mutation C/G to T/A at position 1476
CC which introduces a TGA codon and results in premature termination
CC of the SDD1 polypeptide.
XX
SQ Sequence 2328 BP; 635 A; 487 C; 577 G; 629 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 1,73e-311 Length: 2328
Score: 4009.00 Matches: 774
Percent Similarity: 99.87% Conservative: 0
Best Local Similarity: 99.87% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 21 Gaps: 0

```

US-09-806-767-2 (1-775) x AAZ93809 (1-2328)

Qy 1 MetGluProLysProPhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSer 20

Db 1 ATGGAACCCAAACCTTTCTCTGCAATATCTTCTTATTTCTTCTTCGTC 60

Qy 21 GluIleLeuGlnLysGlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLys 40

Db 61 GAGATCTCTGCAAGAGCAGACTTACATTGTTGAGTTTCTTCTTATTTCTTCTTCGTC 120

Qy 41 ThrPheAlaSerLysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyVal 60

Db 121 ACCTTTGCTCAAGTTTGGTGGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 180

Qy 61 GluGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGlu 80

Db 181 GAAGAAGAGAGGAGAGGCTTCTTCTGACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 240

Qy 81 GlyPheAlaAlaGlnLeuThrGluSerGluAlaGluLeuLeuArgTyrSerProGluVal 100

Db 241 GGATTTGCTGCTCAGTTGACTGAATCAGAAAGCCGAGATACCTGAGATATTCACCTCAAGTT 300
Qy 101 ValAlaValArgProAspHisValLeuGlnValGlnThrThrTySerTyLysPheLeu 120
Db 301 GTTGCAAGTGAAGCCTGACCATGTTCTTCAGGTTCAAAACCATCTACTCTTACAAGTCTTG 360
Qy 121 GlyLeuAspGlyPheGlyAsnSerGlyValTyrSerLysSerArgPheGlyGlnGlyThr 140
Db 361 GGCCTCGACGGTTTGGAAACTCCGGTGATGGTCTAATCTCGGTTGGTCAAGGCACA 420
Qy 141 IleIleGlyValLeuAspThrGlyValTyrProGluSerProSerPheAspAspThrGly 160
Db 421 ATTATCGCGGTGCTTGATCTAGTCTGAGTTTGGCCCTGAAAGTCTTAGCTTTCACGATACCGGA 480
Qy 161 MetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlySerPheSerSer 180
Db 481 ATGCTTCGATTCACCGGAAATGGAAGGATTTGCCAAGAAGGAGAAATTCAGTTCT 540
Qy 181 SerSerCysAsnArgLysIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200
Db 541 TCAGCTGTAAACCGGAAGCTAAATCGGTGCTAGATTCTTTCATCAGAGACACCGTGTGCT 600
Qy 201 AsnSerProGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThr 220
Db 601 AATTCACGAGGAATCACCAACATGCTCGTGNATACATTTCCGCAAGAGATTCAACG 660
Qy 221 GlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaVal 240
Db 661 GGACACGGGACTCACACCGGCTCAACAGTTGGTGGATCCTCTGTTTCGATGCGAATGTT 720
Qy 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyr 260
Db 721 CTTGGCAATGGAGCTGGTGTGGCTGGTGGATGGCTCTGGAGCTCACATTCAGTCTAT 780
Qy 261 LysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaAlaIleAspValAla 280
Db 781 AAAGCTGTGGTTCAATGGTTGTTACAGCTCTGCATCTACAGCATATAGATAGAGCG 840
Qy 281 IleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyr 300
Db 841 ATTCAGATAAAGTCGATGTTCTTTCGCTTTCCTTGGCGGTTTCCTATTCCTTTGTAT 900
Qy 301 AspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCys 320
Db 901 GATGACACAATCGCATTTGGAACATTCGAGCCATGGAACCGGTATATCTGTAATCTGT 960
Qy 321 AlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTyrValSer 340
Db 961 GCAGCTGGTAACAACCGGTCCATCGAAAGCTCTGTTCGAAACACAGCTCTTGGGCTCA 1020
Qy 341 ThrIleGlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGly 360
Db 1021 ACCATTGGCGAGCAGCAGCTGTATCGAAGATTTCCTGCTGGTTCAGATTAGCAACGGA 1080
Qy 361 LysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGlu 380
Db 1081 AAGCTTCTCTATGAGAGTCAATCTATCCGGAAAGGATTAAGAATGCGCGGAGAG 1140
Qy 381 ValGluValIleTyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySer 400
Db 1141 GTTGAGGTGATTATCGTCACAGGAGGAGATAAGGAAGTAGTGATCTCTTTGAGAGGTCA 1200
Qy 401 LeuProArgGluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArg 420
Db 1201 CTTTCCAGAGAAGAAATCCGAGGCAAAATGGTGATTTGTGATTCGGGAGTCAATGGAGA 1260
Qy 421 SerGluLysGlyGluAlaValLysGluAlaGlyValAlaMetIleLeuAlaAsnThr 440
Db 1261 TCGGAGAAAGGAGAGCGGTTAAAGAAGCTCGAGGAGTTGCAATGATCTTAGCCAAATACA 1320
Qy 441 GluIleAsnGlnGluAspSerIleAspValHisIleLeuLeuProAlaThrLeuIleGly 460

Db 1321 GAGATCAACCAAGAAAGATTCATTGACGTTTCATCTCTTACCAGCTACATTGATGCT 1380
Qy 461 TyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArg 480
Db 1381 TACACTGAGTCACTCTCTCTGAAGGCTTATGTTAATGCCACGCTGAACCAAGGCGCG 1440
Qy 481 IleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSer 500
Db 1441 ATAATTTTGGTGTACGCTGATTTGGAGGTCATGAGCACCGGAGGTGCTCAGTTTCA 1500
Qy 501 AlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGly 520
Db 1501 GCTCGAGGACCGAGTTAGCCAAATCTTCGATCTAATAACCGGATGATGATGCTCGGGA 1560
Qy 521 ValAsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSer 540
Db 1561 GTCNATATCATTTGGGCTTGGCCCTCAAAATCTAGGACCAACCGACTTCCTTATGATCA 1620
Qy 541 ArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGly 560
Db 1621 AGAAGAGTTAACTTCACTGTAATGTCAGGAACCTCAATGTCTTGTCCACATGTTAGCGGA 1680
Qy 561 IleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAla 580
Db 1681 ATCACTCTCTTATCCCGTCTGCATACCCGAATGGTCTCCAGCTGCAATCAATCCGCA 1740
Qy 581 LeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLys 600
Db 1741 TTGATGACACAGCGGATTTGTACGATCTCAAGGGAACGATAAAGATGTTAACA 1800
Qy 601 ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 1801 CCACCGGTGTGTTGGATTTGGAGCAGGCGCATGTGAATCCGCAAAAGGCGATAAACC 1860
Qy 621 GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPhe 640
Db 1861 GGATTTGGTTTACACATTCACACAGTGGATTTACATAACTTACCCTCTGCATCTTGGATTC 1920
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 1921 ACAAAGTCAGATATTTTAGCAATCACTCATAGAACGTCGAGTCGCAATGGANATTCGGG 1980
Qy 661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 1981 AAAAACCCGGGTTTGTAGTCTCAATACCGTCGATGCGGTGATTTTCAAACGTTGCAAG 2040
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSer 700
Db 2041 ACTACGAGATGATCACAAAGCGTGTCACTAACGTTGGGAGTCTCTAACTCGATATCTCA 2100
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db 2101 GTGAATGTCAAGGCTCCAGAGGGGATCAAGATTTATTGTCAATCTTAAGAGACTTGTGTT 2160
Qy 721 LysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuLysLysLysAsnArg 740
Db 2161 AAACACGTGGATCAGACGCTGAGTATAGATATGGTTTGTATTGAAGAAGAAACAGA 2220
Qy 741 GlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeu 760
Db 2221 GGAGGAAGGTGGCTAGCTTTTGCAAGGCGAGTTGACTTGGGTCAACTCTCATAATCTG 2280
Qy 761 MetGlnArgValArgSerProLysSerValThrLeuLysThrAsn 775
Db 2281 ATGACGCGAGTTAGAAGTCCAATCTCTGTAACTTGAAGACTAAC 2325

RESULT 3

AZ93814
ID AAZ93814 standard; DNA; 3140 BP.

XX AC AAZ93814;

XX AC AAZ93814;
DT 16-AUG-2000 (first entry)

```

XX Subtilase (St_P2Sca) of Solanum tuberosum.
DE
DE SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomachs; sugar; water; protein; CO_2; H_2O; CO2; H2O;
KW crop protection; feed; foodstuffs; ss.
XX
XX Solanum tuberosum.
OS
XX
XX Key Location/Qualifiers
FH 1..2301
CDS /*tag= a
    /product= Subtilase St_P2Sca
W0200022144-A2.
XX
XX 20-APR-2000.
XX
XX 12-OCT-1999; 99WO-EP07633.
XX
XX 12-OCT-1998; 98EP-0119244.
XX
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
XX Berger D, Altman T;
XX
XX WPI; 2000-317995/27.
XX P-ESDB; AAY83304.
XX
XX Novel recombinant DNA molecules encoding subtilisin-like serine
XX protease, useful for producing transgenic plants with altered stomach,
XX lower water consumption and enhanced diseased resistance
XX
XX Claim 1; Page 85-88; 101pp; English.
XX
XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
XX used to produce transgenic plants with altered stomach
XX characteristics. These plants exhibit improved freshness,
XX increased dry weight, reduced leaf temperatures, reduced water loss
XX and lower water consumption and for enhancing the sugar and/or
XX protein content of plant leaves, modulating CO_2 uptake into and H_2O
XX release from leaves, for sustained photosynthesis under high
XX intensity conditions or for the improvement of disease resistance
XX of plants. The transgenic plants and cells of such plants are useful
XX in the preparation of feed, food or additives. This subfamily of
XX Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
XX (See AAZ93808).
XX
XX SQ Sequence 3140 BP; 939 A; 627 C; 690 G; 882 T; 2 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,49e-213 Length: 3140
XX Score: 2786.50 Matches: 524
XX Percent Similarity: 82.48% Conservative: 107
XX Best Local Similarity: 68.50% Mismatches: 125
XX Query Match: 69.35% Indels: 9
XX DB: 21 Gaps: 5
XX
XX US-09-806-767-2 (1-775) x AAZ93814 (1-3140)
XX
XX 7 PheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLysGln 26
XX
XX 16 TTTCTATGCTTTTACTATGTTTGTTCATACAGCTCAAGATTG-----CAA 66
XX
XX 27 ThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPhe 46
XX
XX 67 ACTTACATAGTTCAGTTCATCCATCGAGCAACAGACCCCTTTTACTCTTAAACTA 126
XX
XX 47 AspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 66
XX
XX 127 CAATGGCACCTTCTTCTTCGCAAAAGCA-----GTTCTCTGGAGAACAGAC 177

```

```

QY 67 ProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeu 86
DB 178 TCGCTTCTCGTCTTTTGTACTCTTACCATCTCCGATGGAAGGTTTGGACGTCGACTC 237
QY 87 ThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValAlaProAsp 106
DB 238 ACTGAAGATGAGGTGAGTTGTTAAGGAATCTAATGATGTGTGTGATACGTCGCTGAG 297
QY 107 HisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGly 126
DB 298 AGGAGGCTTGAATTCAGACTACTTATTCITACAAGTCTTGGGATTAAAGTCCA--ACG 354
QY 127 AsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAsp 146
DB 355 AGAGAAGGAGCTTGTTCAGAGTCTGGATTTCGTCAGGGGCGATCAATGGAGTGTGGAT 414
QY 147 ThrGlyValTrpProGluSerProSerPheAspThrGlyMetProSerIleProArg 166
DB 415 ACTGGAGTTTGGCCAGAAAGTCCAAGTTTTCATGATCATGGGATGCCACCTGTCACAG 474
QY 167 LysTrpLysGlyIleCysGlnGluGlySerPheSerSerSerSerSerCysAsnArgLys 186
DB 475 AAGTGGAGGGTGTCTGCCAAGGAGGACAGGATTTTAATCTTCTAGTTGTAATCGCAAG 534
QY 187 LeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAlaAsnSerProGluGluSer 206
DB 535 CTTATTGTGCAAGGTTTTTTCAGAAAGGACATCGTGTGGCT---TCAATGACATCATCA 591
QY 207 ProAsnMetProArgGluTyrIleSerAlaArgSerThrGlyHisGlyThrHisThr 226
DB 592 CCAGATGCAAGTGGAGGAATATGTCTGCCACGGGATTCCTCATGGCCATGGTACACATACA 651
QY 227 AlaSerThrValGlyGlySerSerValSerMetAlaAlaValLeuGlyAsnGlyValagly 246
DB 652 GCATCCACTGCTGGAGGAGCTGCATTCCTTCCTGCTGCTCGGAAATGGAGCAGGG 711
QY 247 ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsn 266
DB 712 GAGGCTCAGGATGGCCCGGGTGGCCACATTCGAATATATAGTATGCTGTTTCACT 771
QY 267 GlyCysTyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysValAsp 286
DB 772 GGTGTTACAGCTCTGATATACCTTCAGCAATGGATGTGGCCATCAGAGATGGAGTAGAC 831
QY 287 ValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIle 306
DB 832 ATATTGTCACTCTCATCTGGTGGCTTCCTATTCCACTTATGATGATGATCTATTGCGCAT 891
QY 307 GlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGly 326
DB 892 GGAAGTTTCCGAGCCATGAGCATGGAAATTTTCAGTTATATGTGCTGCGGGAATATGGA 951
QY 327 ProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyValaglyThr 346
DB 952 CCAATCCAAAGTTCAGTAGCCACCGTGTCTCTTGGATTGGCCACTATTGTTGCTAGCACA 1011
QY 347 LeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGlu 366
DB 1012 CTTGACAGGAGATTTCCAGCGTTCAGTTTCAGTAGGCAACGGAAAGTTCTCTACGAGAA 1071
QY 367 SerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrVal 386
DB 1072 TCTTGTGATCCCTGGGAAGAAATTCCTAGCTCTCAGAAAGATCTTGAGATCGTTTATGTA 1131
QY 387 ThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIle 406
DB 1132 AAGGATAGGACAGGAGGAGTGAATTTTGTCTTGAGAGGATCGCTATCAAAGCACAGTC 1191
QY 407 ArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAla 426
DB 1192 CGAGGGGAAATGTTGTGTGTGATAGGGAGTCAATGGAGGCGAGAGAAAGGCGAGGTT 1251
QY 427 ValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGlu 446

```

```

Db      1252 GTGAAGGAGGAGTGGTCTCCATGATCTTAGCAATACAGCAATAATATGAGGAA 1311
Qy      447 AspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyThrThrGluSerValLeu 466
Db      1312 GATTCCATTGATCTCCATCTCTCCAGCAAGCTTGATTGGCTTCGATGAATCAATCAA 1371
Qy      467 LeuLeuAlaThrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyThr 486
Db      1372 TTCAAAATACCTGATCACTCAAAAGAGCAACACAGCTCGATTATATTTGGAGGACG 1431
Qy      487 ValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeu 506
Db      1432 GTAATAGGAAGCTAGACACCTCGTAGCTAGCTAGTTCTCAAGGGGGCCAGCTAT 1491
Qy      507 AlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAla 526
Db      1492 ACTGATCCTTTCAATTCTCAAACTGATTGATTGCTCCAGGGGTAACATAATTCGGCT 1551
Qy      527 TrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThr 546
Db      1552 TGGCCCAAAACTAGGCCCTGCTCTCCGAAAGATTTCAGAGAGTAATTCAT 1611
Qy      547 ValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArg 566
Db      1612 GTTATGTCAGGACCTCAATGGCATGTCTCTCAATGTAAGTGGTAATGCCCATGTCTCAT 1671
Qy      567 SerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAsp 586
Db      1672 TCAGCTCATCTTAATGGACTCCAGCAGCAATAAGATCCGATTAAATGACCACTGCAGAT 1731
Qy      587 LeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAlaGlyValPheAla 606
Db      1732 ACAGCTGATCATATGGGAAACCAATCATGGATGGAGATGCACCACTAAACTTTTGCA 1791
Qy      607 IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIle 626
Db      1792 GCTGGAGCTGGACACGTAACCTCGGAAGCCATGCTGATGATATATATGATC 1851
Qy      627 GlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeu 646
Db      1852 CAGTTGATGAATATATCATCTATCTTTGACATATCGGATACAGAAATTCGAGTCTTC 1911
Qy      647 AlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSer 666
Db      1912 AGCATTAATCATAGGAATGTCAGTGCATGATCATTTACAGAACCAACAGGGGTTTCAGC 1971
Qy      667 LeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThr 686
Db      1972 CTAAATTTACCCCTCAATTTCAATTAACCTTCAGAGCAGGAATGATAGAAAGATAATCAAG 2031
Qy      687 ArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaPro 706
Db      2032 AGGAGAGTAACAAATGTGGGAAACCCCTTAATCTATTTTACTCAGTTGACATGAGGACCT 2091
Qy      707 GluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThr 726
Db      2092 GAGGAGTCAAAGTGAGAGTGAAACCCAGCTGCTGATATTTAAACATGTGACCAAGC 2151
Qy      727 LeuSerTyrArgValTppPheValLeuLysLysAsnArgGlyLysValAlaSer 746
Db      2152 TTAAGCTATAGAGTTGGTTTATA--TCACGAAAGAAATAGAGTCTTAAAGGATCAGC 2208
Qy      747 PheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSer 766
Db      2209 TTTGCAGAGGGGCAATTGACATGTTTCAATGTAGGAAACAAAGCCAGAAAGTTAAAGT 2268
Qy      767 ProfileSerValThr 771
Db      2269 CCTATTTCCTGCACA 2283

```

RESULT 4
AAZ93813

```

ID      AAZ93813 standard; DNA; 2492 BP.
XX      AC
XX      AAZ93813;
XX      16-AUG-2000 (first entry)
XX      Subtilase (Pigesp) of Solanum tuberosum.
XX      SDD1; serine protease; subtilisin; transgenic plants; dry weight;
XX      stomach; sugar; water; protein; CO_2; H_2O; CO2; H2O;
XX      crop protection; feed; foodstuffs; ss.
XX      Solanum tuberosum.
XX      Key Location/Qualifiers
XX      CDS 3..2492
XX      /*tag= a
XX      /product= Subtilase Pigesp
XX      WO200022144-A2.
XX      PD 20-APR-2000.
XX      PF 12-OCT-1999; 99WO-EF07633.
XX      PR 12-OCT-1998; 98EP-0119244.
XX      (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX      (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX      PI Berger D, Altmann T;
XX      WPI; 2000-317995/27.
XX      P-PSDB; AAY83303.
XX      Novel recombinant DNA molecules encoding subtilisin-like serine
XX      protease, useful for producing transgenic plants with altered stomata,
XX      lower water consumption and enhanced diseased resistance
XX      Claim 1; Page 79-82; 101pp; English.
XX      Sequences encoding SDD1, a subtilisin-like serine protease, can be
XX      used to produce transgenic plants with altered stomata
XX      characteristics. These plants exhibit improved freshnees,
XX      increased dry weight, reduced leaf temperatures, reduced water loss
XX      and lower water consumption and for enhancing the sugar and/or
XX      protein content of plant leaves, modulating CO2 uptake into and H_2O
XX      release from leaves, for sustained photosynthesis under high
XX      intensity conditions or for the improvement of disease resistance
XX      of plants. The transgenic plants and cells of such plants are useful
XX      in the preparation of feed, food or additives. This subtilease of
XX      Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
XX      (See AAZ93808).
XX      SQ Sequence 2492 BP; 766 A; 480 C; 511 G; 734 T; 1 other;

```

Alignment Scores:
Pred. No.: 9,74e-209 Length: 2492
Score: 2727,50 Matches: 521
Percent Similarity: 79,22% Conservative: 108
Best Local Similarity: 65,62% Mismatches: 132
Query Match: 67,88% Indels: 33
DB: 21 Gaps: 5

US-09-806-767-2 (1-775) x AAZ93813 (1-2492)

Qy 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLys 25
Db 102 TATTTCTCTGT-----TTTCTACTCTGTTTATCCCTGCTACAGCTCAAAATTTG 155
Qy 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThr---PheAlaSer 44
Db 156 CAAACTTATATAGTACAAATATACATCCACAAATGATCATCAACAGAACCCCTTTAGTTCT 215

Db 2370 AGGAGAGATTGGGACTCAAGAGAGAGCTTTGCAGAGGACAAATTCATGTGGATCAAC 2429
QY 757 SerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
Db 2430 TCCAGAGATAAATACAGAAAGTTAGAAGTCTCTATTTCAGTT 2471

RESULT 5

AAZ93812
ID AAZ93812 standard; DNA; 3865 BP.

XX AC AAZ93812;

DT 16-AUG-2000 (first entry)

XX DE Subtilase (Plgsp) of Solanum tuberosum.

XX KW SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
KW crop protection; feed; foodstuffs; ss.

XX OS Solanum tuberosum.

XX FH Location/Qualifiers

FT exon

FT 3...551

FT /*tag= a

FT /label= Exon 1

FT intron

FT 552...966

FT /*tag= b

FT /label= Intron 1

FT exon

FT 967...1654

FT /*tag= c

FT /label= Exon 2

FT intron

FT 1655...1737

FT /*tag= d

FT /label= Intron 2

FT exon

FT 1738...2222

FT /*tag= e

FT /label= Exon 3

FT intron

FT 2223...2485

FT /*tag= f

FT /label= Intron 3

FT exon

FT 2486...3252

FT /*tag= g

FT /label= Exon 4

XX WC2000022144-A2.

XX 20-APR-2000.

XX 12-OCT-1999; 99WO-EP07633.

XX 12-OCT-1998; 98EP-0119244.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

XX Berger D, Altmann T;

XX WPI; 2000-317995/27.

XX P-PSDB; AAY83302.

XX Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
FT lower water consumption and enhanced diseased resistance

XX Claim 1; Page 72-76; 101pp; English.

XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or

CC protein content of plant leaves, modulating CO₂ uptake into and H₂O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This subtilease of
CC Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
CC (See AAZ93808).

XX SQ Sequence 3865 BP; 1202 A; 707 C; 715 G; 1237 T; 4 other;

Alignment Scores:

Pred. No.: 9,06e-193 Length: 3865
Score: 2531.50 Matches: 524
Percent Similarity: 60.40% Conservative: 109
Best Local Similarity: 50.00% Mismatches: 129
Query Match: 63.00% Indels: 288
DB: 21 Gaps: 9

US-09-806-767-2 (1-775) x AAZ93812 (1-3865)

QY 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLys 25

Db 102 TATTTTCTCTGT-----TTTCTCTCTGTCTTTTATTCCTCTCAAGCTCAAAATTG 155

QY 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThr---PheAlaSer 44

Db 156 CAACTTATATAGTACAAATACATCCACACATGCATCAACAAGAACCCCTTTAGTTCT 215

QY 45 LysPheAspTrpHisLeuSerPheLeuGlnGlu----- 55

Db 216 AATTTTCAGTGGCACCTTTTCATTTCTTGAAATTTTCAAAACATTCATTATTCACATTC 275

QY 56 -----Ala 56

Db 276 AAATATATTCATGGAACCAATCCAAATCTATATTTCTGTCTTCTGTATTATTC 335

QY 57 ValLeuGlyValGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGly 76

Db 336 CTGCTACAGCATTTTCCCTCAGTGAAACTCGAGTTCTCGCTTTGTTACTCTTACCAT 395

QY 77 SerAlaIleGluGlyPheAlaIleGlnLeuThrGluSerGluAlaGluLeuLeuArgTyr 96

Db 396 TCTGCAATTGAAGGTTTTCAGCACCTTCTATCTGAAATGAGCTAAAGGCACCTGAAGAAA 455

QY 97 SerProGluValValAlaValArgProAspHisValLeuGlnValGlnThrTyrSer 116

Db 456 TCGAATAATGTGTATCAATATATATCCGAGAGAGAGCTTGAGTTCAAACTATTCT 515

QY 117 TyrIysPheLeuGlyLeuAsp----- 123

Db 516 TACAAGTTCTTAGGACTTAGTCTCTCAAGGAAGGTATGTTACATAGTTTGTATATATA 575

QY 123 ----- 123

Db 576 TAAAGTTGAGGAACAATGAGTCATAGGCTTTTATTATTAAATGACTAGTCTGGACTAG 635

QY 123 ----- 123

Db 636 AGTCTCACGCTAGGATTTAAACTTTGTGAATTAAGGAAAATTTCTTAAACCCCTTTACAA 695

QY 124 -----Gly 124

Db 696 CTGCTCTGGAGTCATTTCTTATGTCAATGTGATTCACAGTTTCATATATTAACAAAAGGA 755

QY 125 Phe-----GlyAsnSer----- 128

Db 756 TTGTCTTTACCTTGTGTTTCGAAGTAAATAATTTCAACGAGGGGATTCACCGAACCC 815

QY 128 ----- 128

Db 816 CATGGTTCCCTCAACTCTGCTAATAGATATTTTCTATGTATTTCATTATTAAGACTTGAA 875

QY 128 ----- 128

Db 876 TTTTAAAGCTTTAAATTTCTATGTTCCCTCGGAGCCTTTCTNCTNACTTACTTTTATAC 935
Qy 129 -----GlyValTTPSerLysSerArgPheGlyValnG1 139
Db 936 TGTCTTTGATCTTTTTTTTCTTAATAAGGTACTTGGTTAAAGCTCTGGATTTGGTCGAG 995
Qy 139 yThrIleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspTh 159
Db 996 CGCGATCATTTGGAGTCTTTGATACTGGAATTTGGCCAGAAAGTCCAAGTTTGTGATCA 1055
Qy 159 rGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSe 179
Db 1056 TGGAAATGTCCTCTATTCCAAAGAAATGGAAGGTTCTGCCAAGAAAGAAACCTCAA 1115
Qy 179 rSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgVa 199
Db 1116 TTCTTCAAGTTGCAATCGCAAGCTTTATGTCGCAAGGTTTTTCCAGATAGGACACATGAT 1175
Qy 199 lAlaAsnSerProGluGluSerProAsnMetProArgGluTyIleSerAlaArgSpSe 219
Db 1176 GGCATCAAGACATCAAAATCAATAGATTTTATGGAGGATTAATGATACCTCGAGATTC 1235
Qy 219 rThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAs 239
Db 1236 TCAAGGCCATGCTTACATACATACAGCATCTACTGCGAGGGGAGCTCCCGTTCCAATGGCGAG 1295
Qy 239 nValLeuGlyAsnGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaVa 259
Db 1296 TGTGCTTGGAAATGGAGCAGAGAGGCTCGAGGGATGGCCCTCGTCAATCGCGAT 1355
Qy 259 lTyIleValLysTrpPheAsnGlyCysTyIleSerSerAspIleLeuAlaIleAspVa 279
Db 1356 ATCAAAAGTTTGTGCTAGTGGTTGTTATAGTTCTGATATATCTTGAGCAATGGATGT 1415
Qy 279 lAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLe 299
Db 1416 AGCTATTAGAGATGGAGTAGACATATGCTCTTTTCAATGGTGGTTTCCCTGTTCCACT 1475
Qy 299 uTyIleAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValI1 319
Db 1476 TTATGAGCACTATTGCTATTGGCAGTTTTCAGCTATGGAACGTGGAATTTCACTTAT 1535
Qy 319 eCysAlaIleGlyAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpVa 339
Db 1536 ATGTCGTGCGAGAAATTAATGGTCAATCTTAAGTTCAGTAGCAAAATGAGGCTCTTGGAT 1595
Qy 339 lSerThrIleGlyValaglyThrLeuAspArgPheProAlaValValArgLeuAla-- 358
Db 1596 TGCACATTTGGTGTAGCACACTTGACAGAAATTTCCAGCAATTAATTCAGCTAGG-TA 1654
Qy 358 ----- 358
Db 1655 TGTACATTTTGTCTTAAATGATATTTCCGGTGTTCACGCTAAATATGTGTCCT 1714
Qy 359 -----AsnGlyLysLeuLeuTyIleGlySerLeuTyIleProG1 371
Db 1715 TCATTCATATTTTCCACAGGTAAATGGCAAGTATGTGTATGGAGAATCTTGTACCCGG 1774
Qy 371 yLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyIleValThrGlyGlyAspLy 391
Db 1775 CAACAAGTTCAATATCTCAGAAAGTTCTTGAGATTTGTTATCTCAATGACGGTGAATA 1834
Qy 391 sGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetVa 411
Db 1835 TGGNAGTGAATTTTGTATAGGGTCTCTGCCAAGAGCTAAAGTCCATGGAAATAATCGT 1894
Qy 411 lIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaG1 431
Db 1895 TGTATGTGATCGTGGAGTTAATGAAGACAGAGAGAAAGGTCAAGTTGTTAAAGAAACAG 1954
Qy 431 yGlyValAlaMetIleLeuAlaSerThrGluIleAsnGlnGluAspSerIleAspVa 451

```
QY 723 lAspGlnThrLeuSerTyrArgValTrpPheValLeuLysLysLysAsnArgGlyGly 743
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3094 TAATCAAGTTTAAGTTACAGAGTTTGGTTTATA---TCAAGGAAGAGATTGGACTCA 3150
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 743 sValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnAr 763
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3151 AAGGAGAAGCTTTGCAGAGGACCAATTGATGTTGGATCAACTCCAGAGATAAATACCAGAA 3210
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 763 sValArgSerProIleSerVal 770
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3211 AGTTAGAGTCCTATTTCAGTT 3232
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
AAC43317
ID AAC43317 standard; DNA; 2265 BP.
XX
AC AAC43317;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 38825.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127452.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
```


943 AGAGCTTCTGTTGCCAATGTTGCTCCTCGGTTATGACTGTTGGTCTGCTATCTTAGAT 1002
Qy
349 AtcArgPheProAlaValValArgLeuAlaAsnGlyValLeuLeuTyrcGlyGluSerLeu 368
Db
1003 AGAGATTTTCGGGCTTTTCGGAATCTCGGTAAACCGGAACGACTTACCGGTGTTTCGCTG 1062
Qy
369 TyrProGlyLysGlyLysAsnAlaGlyArgGluValGluValileTyrcValThrGly 388
Db
1063 TATAGCGGTGAGGAATG-----GGGACGAAGCCGCTTGAATGGTTTATATAAAGG 1116
Qy
389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluLeuArgGly 408
Db
1117 AATAGTAGTTCGAGTAATCTTTCTTACCTGGTTCGCTTCGATTCCGAGTATTGTTGGTGG 1176
Qy
409 LysMetValLysCysAspArgGlyValAsnGlyArgSerClyLysGlyGluAlaValLys 428
Db
1177 AAGATTCTGTTGTAGTAGAGTGTAAATGCTAGAGTTGAGAAAGAGAGCTGGTGTAGA 1236
Qy
429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSer 448
Db
1237 GATCTGCTGGTGTAGGATGATATGCGGAATACTGCTCGAGTCGAGAGGAGCTGTG 1296
Qy
449 IleAspValHisLeuLeuProAlaThrLeuIleGlyTyrcThrGluSerValLeuLeuLys 468
Db
1297 GCGGATAGTCTATTTGCTTCGCGGATCGCTGTAGGGAAGAGACTGGTGATTTACTTAGG 1356
Qy
469 AlaTyrcValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIle 488
Db
1357 GAGTATGTTAAGTCAGATCTTAAACCAACCGCTCTCTCTGTTTAAAGGAACGCTTCT 1416
Qy
489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508
Db
1417 GACGTTAAGCGCTCTCTGTTGGTCTGCTTTAGCTCGAGAGGCTCTAATCTGTTACT 1476
Qy
509 ProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaIleAlaIlePro 528
Db
1477 CCTGAATCTTGAGCCCTGATGTTATTGGTCTCGAGTAAATTTTGGCTGGTGTCT 1536
Qy
529 GlnAsnLeuGlyProThrGlyLeuProTyrcAspSerArgArgValAsnPheThrValMet 548
Db
1537 GACGCTATTGCTCTACTGCTCTTGACAAGGACTCTAGGAGGACTCAGTTTCAACATCATG 1596
Qy
549 SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAla 568
Db
1597 TCAGGTACGTCATGTCATGCCACATCAGTGGTTTACGGGCTCTTTTGAAGCAGCT 1656
Qy
569 TyrProAsnTyrSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrc 588
Db
1657 CACCTGAGTGGATCCGAGTGGCTATCAATCAGCTCTCATGACTACAGCTTACGTTCTT 1716
Qy
589 AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606
Db
1717 GACAACACCAACGCTCTCTCCATGATGCTGCAGACACACAGCTATCTAACCCATATGCT 1776
Qy
607 IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrcAsnIle 626
Db
1777 CACGGCTCGGCGCATGTAGATCCCAAGGCTCTCTCACCAGGCTCTGTGCTACGATC 1836
Qy
627 GlnProValAspTyrcIleThrTyrcLeuCysThrLeuGlyPheThrArgSerAspIleLeu 646
Db
1837 TCACCGAGGATACATCAGTTTGTGGCTCTCTAGACTACAGCTGCTGATCATGTTT 1896
Qy
647 AlaIleThrHisLys---AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe 665
Db
1897 GCGATTGTGAAGCGACCTAGCGTTAACTGCTCGAAGAGTTC---TCAGATCCCTGGT--- 1950
Qy
666 SerLeuAsnTyrcProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIle 685
Db
1951 CAGCTCAACTACCAAGTTTCTCGGTTTGTGTT---GGGGGTAAGAGAGTTGTCGGTAC 2007
Qy
686 ThrArgArgValThrAsnValGlySerProAsnSerIleTyrcSerValAsnValLysAla 705
Db
2008 ACTCGGGAAGTAACAAATGTTGGTGCAGCAAGCTCGGTTTACAAGTACGCGTTTATGGA 2067

Qy 706 ProGluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGln 725
Db 2068 GCTCCTAGTGTGCGAATCTCTGTTAAACCATCGAAACTTTCGTTTAAAGCGTGGAGAG 2127
Qy 726 ThrLeuSerTyrcArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAla 745
Db 2128 AAGAAGAGGTACACAGTCACGTTTGTAGCAAGAAAGGAGTGTAGTATGACGAACAAGGCT 2187
Qy 746 SerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GTTTCATCCTTGGAGCAATCCGACGAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTTGCA 2244
RESULT 7
ABZ13260
ID ABZ13260 standard; DNA; 2265 BP.
XX
AC ABZ13260;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1065.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JUN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
PT
XX
PS Claim 144; SEQ ID NO 1065; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 2265 BP; 519 A; 510 C; 542 G; 694 T; 0 other;

Alignment Scores: 8.28e-130 Length: 2265
Pred. No.:

Score: 1742.00 Matches: 370
Percent Similarity: 63.48% Conservative: 115
Best Local Similarity: 48.43% Mismatches: 241
Query Match: 43.35% Indels: 38
DB: 24 Gaps: 15

US-09-806-767-2 (1-775) x ABZ13260 (1-2265)

QY	11	IlePheLeuLeuPheCysSerSerSerSerGluLeuLeuGlnLysGlnThrTrpIleVal	30
DB	52	CTCTCTCTCTCTCCACACACCGCC-----AAAAAACCTACATCATC	96
QY	31	GlnLeuHiProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeu	50
DB	97	CGCGTCAAT---CACTCCGATAAACCGGAATCTCTCACTCACACGAGATGGTACACA	153
QY	51	SerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSerArg	70
DB	154	TCTCAACTCAATTCA-----GAAATCATCT	177
QY	71	LeuLeuTyxSerTyrglySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGlu	90
DB	178	CTTCTCTACACITACACACACCTCTCCATGGCTTCCTCGGTACCTCGACTCCACCGAA	237
QY	91	AlaGlu---IleLeuArgTyrSerProGluValValAlaValArgProAspHisValLeu	109
DB	238	GCGGATTCCTCTCTCTCTCTCTCAAACTCAATCTCGATATCTTCGAAGATCTCTCTCTAC	297
QY	110	GlnValGlnThrThyTyrSerTyrlsPheLeuGlyLeuAspGly---PheGlyAsnSer	128
DB	298	ACACTTCACACTACCGGTACTCCTGAGTTTCTCGGTCTCAATTCGGAATTCGGTGTTCAC	357
QY	129	GlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGly	148
DB	358	GATCTCGGTCTCTCTCT-----AACGGCTTATCATCGGAGTTTAGATACTGGC	408
QY	149	ValTrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrp	168
DB	409	GTTTGGCCTGAATCTAGAAAGCTTCGATGATCTCATATGCTCGAGATCTCTTCTAAATGG	468
QY	169	LysGlyLysCysGlnGluGlySerPheSerSerSerCysAsnArgLysLeuIle	188
DB	469	AAAGGAGATGTAATCTGGTTCCGATTCGATTCGAATGTTGTGTAAAGAGAGCTTATC	528
QY	189	GlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsn	208
DB	529	GGAGCTAGAGAGCTTCTCCAAAGAGATTCAAATGCTCTCTGGTGGTTCGAGTAAG	588
QY	209	MetProArgGluTyrlsSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer	228
DB	589	-----CGTGAATCTGTTTCTCTCTCGTATGTTGACGGACATGGAACACATACTTCAACT	642
QY	229	ThrValGlyGlySerSerValSerMetAlaValLeuGlyAsnGlyAlaGlyValAla	248
DB	643	ACCGCGCGGGATCCGCGTTTGAACACGCTAGCTTCCTCGGTATACGCCCGGTACGGCC	702
QY	249	ArgGlyMetAlaProGlyAlaHisIleAlaValTyrlsValCysTrpPheAsnGlyCys	268
DB	703	AGAGGTATGCCCACTCGTGCTCGTGCTGCTACTATAAAGTTGTTGCGACTCTGGTTGT	762
QY	269	TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeu	288
DB	763	TTTGGATCTCATATACCTAGCTCTATGGATCGAGCTATACTTGTATGTTGTATGTCCTT	822
QY	289	SerLeuSerLeuGlyGlyPheProIleProLeuTyrlsAspThrIleAlaIleGlyThr	308
DB	823	TCGTTATCTCTTGGTGGTGGTCTCTCGGTATTATCGCGATACGATTTGGATTTGGAGCG	882
QY	309	PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle	328
DB	883	TTTTCGGCTATGAGAGAGGTGTTTTGTGCTCTTCTGCTGGTATATAGTGTCCTACT	942
QY	329	GluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAsp	348


```
Db 2008 ACTCGGAAGTAACAATGTTGGTCGACGAAGCTCGTTTACAAAGTGACGGTTAATGGA 2067
Qy 706 ProGluGlyLeuLysValIleValAlaSerProLysArgLeuValPheLysHisValAspGln 725
Db 2068 GTCCTAGTGTGGGATCTCTGTTAAACCATCGAACTTCGTTTAAAGCGTGGAGAG 2127
Qy 726 ThrLeuSerTyrArgValTrpPheValLeuLysLysLysLysLysLysLysLysValAla 745
Db 2128 AAGAAGAGGTACACACTCACGTTTGTACCAAGAAAGGATGAGTATGACGAACAGGCT 2187
Qy 746 SerPheAlaGlnGlyGlnLeuThrTrpValAlaSerHisAlaLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GGTTCCATCACTTGGAGCAATCCGACGAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTTGCA 2244

RESULT 8
AAC46158
ID AAC46158 standard; DNA; 2637 BP.
XX
AC AAC46158;
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49118.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 03-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
```



```

Db      1022 GCTATCGCTATTGGAGCTTTTGGAGCTATTGACAGAGGCATATTCTGCTCTCTCCGCC 1081
Qy      323 GIVASnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTyrValSerThrIle 342
Db      1082 GGTAAACGGAGGTCCTGCTTTAACGGTGACAAATGTTGCTCCGTTGGATGACACAGTC 1141
Qy      343 GlyAlaGlyThrLeuAspArgPheProAlaValAlaValArgLeuAlaAsnGlyLysLeu 362
Db      1142 GGAGCTGGAACAATCGATAGGATTTTCCGGCCAAATGTGAACCTCGGTAACCGGAAGATG 1201
Qy      363 LeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGlu 382
Db      1202 ATTCTCGGTGTAGTGTATACGGTGCACCGGCTCTG---GATCCGGGTGCAATGACCGG 1258
Qy      383 ValIleTyr-----ValThrGlyGlyAsp---LysGlySerGluPheCysLeuArg 398
Db      1259 CTGTGTTACGGTGTAGTGTACTAGCGCGGAGTGTACTCTCTCTGTCTGTCTTGA 1318
Qy      399 GlySerLeuProArgGluGluLeuArgGlyLysMetValIleCysAspArgGlyValAsn 418
Db      1319 GGCTCGTGTGATCCGAATTTAGTGAAGGAAGATCGTTCTTTGTGATAGAGAAATCAAT 1378
Qy      419 GlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAla 438
Db      1379 TCTAGAGCAACCAAGGTGAGATCGTACGGAATAATGGAGGCTTGGGGATGATTAGCG 1438
Qy      439 AsnThrGluLeuAsnGlnGluLeuAspSerIleAspValHisLeuLeuProAlaThrLeu 458
Db      1439 AATGGTGTGTTTACGGTGAAGGTTTGTAGTGTGATTGCCACGTGTACCGCGACATCT 1498
Qy      459 IleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLys----- 476
Db      1499 GTTGGTCTCTCGAGGAGATGAGATTCGTAGTATATCTCTCAATCATCCAAATCTCGT 1558
Qy      477 -----ProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArg 492
Db      1559 TCATCGAAACATCCAAACGGTACAGTGTGTTTCAAAGGTACCGAGTCCGAGCT 1618
Qy      493 AlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeu 512
Db      1619 GCACCGGTTGTCATCTTTCTCTGCTCGTGGTCTTAATCCAGAGACCGCGAGATTCCT 1678
Qy      513 LysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTyrProGlnAsnLeuGly 532
Db      1679 AATCCGATGTATCCGACCTGCTGTTGATATTTCTAGCTGCTTGGCTCACCAGATGCT 1738
Qy      533 ProThrGlyLeuProTyrAspSerArgValAlaAsnPheThrValMetSerGlyThrSer 552
Db      1739 CCATCTGCTGTACTTCTGATACCGGAGAGTGTCAACATTTTATCAGGCACTTCG 1798
Qy      553 MetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTyr 572
Db      1799 ATGCGGTGCGCGACGCTGTCTGTCTAGTGTCTTCTTAAGCGCGCTCATCCGATTGG 1858
Qy      573 SerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGly 592
Db      1859 AGTCCGCGCGGATAGATCAGCATTTGATAACCAACCGCTTACACGGTTGATTAACAGCGGT 1918
Qy      593 LysAlaIleLysAsp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAla 609
Db      1919 GAGCCGATGATGATGATGATGATCCATGCGAAT---ACATCTTCGTTATGATGATGTTG 1975
Qy      610 GlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProVal 629
Db      1976 GGTATGTTTCAACCAACCAAGCTATGATCCGGATTAGTCTATGATATATCAATCTTAT 2035
Qy      630 AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThr 649
Db      2036 GATTATATCAACTCTTGTGTAAATCTTAACCTACACTAGAACCAACATTTGTCAATAACC 2095
Qy      650 HisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyr 669

```

```

Db      2096 CGTCCCAAGCGGACTGTGACGGTCAAGACGAGCGGGACACGTCGGGAATTTAACTAC 2155
Qy      670 ProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThr----- 686
Db      2156 CCGTCGTCCTCTGCTGCTGTTTCAGCAGTATGAGAGAGTAAATATGTCGACACATTTCA 2215
Qy      687 ArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaPro 706
Db      2216 AGACCTGTGCAATATGAGCGATTCCGATTTCGGTTTACGATTAGGATTAGCGCGCCT 2275
Qy      707 GluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThr 726
Db      2276 AGAGGACTACGCTGAGCGTTGAGCGGAGAGCTATCTGTCAGCGGGTGGGCGAGAA 2335
Qy      727 LeuSerTyrArgValTrpPheValLeuLysLysLysAsnArgGlyGlyLysValAlaSer 746
Db      2336 CTGAGT-----TTTGTGTTAGGTTTAAAGCACACAGAGGTGAAGTTGTCAACCA 2383
Qy      747 PheAla-----GlnGlyGlnLeuThrTyrValAsnSerHisAsnLeuMetGln 762
Db      2384 GGAGCGCAAAATCTGGAGACTGCTCATATAGTTTGGTCAGATGG-----AAACGG 2434
Qy      763 ArgValArgSerProIleSerValThrLys 773
Db      2435 AATGTGCAAGTCCCTTGGTGGTCACTCTTCAA 2467

RESULT 9
AAC42012
ID AAC42012 standard; DNA; 2638 BP.
XX
AC AAC42012;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 33960.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
FN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.

```

PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138640.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149910.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.


```
Db 1978 GGTATGTTTACCCACACAGCATATGGATCCGGGACTAGTCTACGATATAACATCTTAT 2037
Qy 630 AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThr 649
Db 2038 GATTATATCACTTCCTGTGTAACTTCACTACACTGGAACCAACATTTGGACGATAACC 2097
Qy 650 HisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyr 669
Db 2098 CGTCCCAAGCGGACTCGACGGTGCAGCGGACGCGGACACGTCGGAATTTGAACCTAC 2157
Qy 670 ProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThr----- 686
Db 2158 CCGTGTCTCTGTCGGTTCACAGTATGGAGAGTAATATGTCGACACATTTCAATT 2217
Qy 687 ArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaPro 706
Db 2218 AGGACTGTGACAAATGTAGCGATTCGGATTCGGTTTACGAGATTAAGATGAGCGGCT 2277
Qy 707 GluGlyIleLysValIleValAsnProLysArgLysValPheLysHisValAspGlnThr 726
Db 2278 AGAGGACTACAGTACAGTGTGACCGGATGACCGGAGAGTATCGTTTCAGCGGTGGGCGAGAA 2337
Qy 727 LeuSerTyrArgValTyrPheValLeuLysLysLysAsnArgGlyGlyLysValAlaSer 746
Db 2338 CTGAGT-----TTTGTGTGTAGGTTAAGACCAACACAGAGTGAAGTTGTCAACA 2385
Qy 747 PheAla-----GlnGlyGlnLeuThrTyrValAsnSerHisAsnLeuMetGln 762
Db 2386 GGAGCGCAAAATGTGGAGACTGGTCATATGTTTGGTCAGATGGG-----AAACGG 2436
Qy 763 ArgValArgSerProIleSerValThrLeuLys 773
Db 2437 AATGTGACAACTCCCTTGGTGTGCACCTTCAA 2469

RESULT 10
AAC47609
ID AAC47609 standard; DNA; 2434 BP.
XX
AC AAC47609;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54454.
XX
KW Hybridisation assay; generic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
```

```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 22-JUL-1999; 99US-0145218.
PR 22-JUL-1999; 99US-0145224.
PR 22-JUL-1999; 99US-0145276.
PR 22-JUL-1999; 99US-0145313.
PR 22-JUL-1999; 99US-0145319.
PR 22-JUL-1999; 99US-0145351.
PR 22-JUL-1999; 99US-0145386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 8.62e-122 Length: 2434
Score: 1642.50 Matches: 342
Percent Similarity: 59.95% Conservative: 128
Best Local Similarity: 43.62% Mismatches: 281
Query Match: 40.88% Indels: 33
DB: 21 Gaps: 11

US-09-806-767-2 (1-775) x AAC47609 (1-2434)

QY 4 LysProPhePheLeuCysIle-----IlePheLeuLeuPheCysSerSer 19
Db 116 AARCCCTTTCTTTTCAATCTTATCAATCAATCTTCTTCTTCAAGCAGAACAACT 175
QY 20 SerGlutLeuGlnLysGlnThrTyrIleValGlnLeuHisProAsnSerGluThrAla 39
Db 176 ACTCAATCTCTACCAAGAGACCTTATGTTATCCACATG---GATAATCTGCCATGCCT 232
QY 40 LysThrPheAlaSerLysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGly 59
Db 233 TTACCTTACACTAATCACTACCTACCTTCTTCTTCAAGATATAACTCTGTAAACCAACAT 292
QY 60 ValGluGluGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIle 79
Db 293 AATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
QY 80 GluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluLeuLeuArgTyrSerProGlu 99
Db 353 CACGGTTAGCAGCTCAGCTTACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412
QY 100 ValValAlaValArgProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPhe 119
Db 413 GTTGTAGCTGTATACCTGAGACAGATACAGAGCTTACACATCAAGAGTCCACAGCTTT 472
QY 120 LeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGly 139
Db 473 CTTGGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 532
QY 140 ThrIleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThr 159
Db 533 GTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 592
QY 160 GlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSer 179
Db 593 GGTATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 652
QY 180 SerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgVal 199
Db 653 AAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACG 712
QY 200 AlaAsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSer 219
Db 713 GCACGGGAGAGATCGATGAA-----GAGCTTGAATATAAGTCAACCGAGAGACAGA 763
QY 220 ThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerValSerMetAlaAsn 239
```


Db	764	GATGGTCACGGGACACACATGCAGCTACTCTAGCTGGCTCACCTGTTAAAGGAGCTAAT	823
Qy	240	ValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaVal	259
Db	824	CTTTTGGATTGCTTATGGACAGCTCGAGGATGGCTCAAAGGCTAGAGTTGCTGCT	883
Qy	260	TyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaIleAspVal	279
Db	884	TATAAGTCTGTTGGTCGGAGGTGTTTTCAGTTTCAGATATTTGTCGGCTGTTGATCAA	943
Qy	280	AlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeu	299
Db	944	GCTGTGCTGATGGAGTTCAGTACTCTCTATATCATTAGGTGGTGGGTCTCTACTTAT	1003
Qy	300	TyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIle	319
Db	1004	TCTAGAGNCAGTTTGTCTATAGCAACATTCGAGCAATGGAGATGGAGATTTCGTTTCG	1063
Qy	320	CysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpVal	339
Db	1064	TGTTCTGCGGTAAATGGAGGTCCTGATCCGATTAGTCTCACTAATGTTTCTCCATGATC	1123
Qy	340	SerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsn	359
Db	1124	ACAAACAGTTGTGCAAGTACTATGATAGAGATTTCAGCAACAGATCAAGATAGGAAC	1183
Qy	360	GlyLysLeuLeuTyrGlyLysLeuTyrProGlyLysGlyIleLysAsnAlaGlyArg	379
Db	1184	ATGAGACATTCAAAGGAGTGTCACTTTACAAAGCAGAACAGTTTTGCTTCAAGATATAA	1243
Qy	380	GluValGluValIleTyrVal-----ThrGlyGlyAspLysGlySerGluPheCys	396
Db	1244	CAGTATCCTCTGGTTTACTTAGGAAGGAATCCAGTAGTCTGATCCACCTCGTCTGT	1303
Qy	397	LeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArgGly	416
Db	1304	CTAGATGGAGCTTTGGATCGCGGCATGTACGGGAAAGATCGTGATATGCACCGCGGT	1363
Qy	417	ValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIle	436
Db	1364	GTTACTCCAGCTGTGCAAAAGGGTCAGTTGTGAAAAGAGCTGTGGAATGGGATGGTT	1423
Qy	437	LeuAlaAsnThrGluIleAsnGlnGluLysAspSerIleAspValHisLeuLeuProAla	456
Db	1424	TTAATAACACTGCAACAATGTTGAGAGGCTTTGTCAGATTCTCATATGCTTCCAGCT	1483
Qy	457	ThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLys	476
Db	1484	GTTGCAGTTGGAGAAAGAGGTAAACTAATCAACAGTACGGCATGACGAGTAAAAAA	1543
Qy	477	ProLysAlaArgIleIlePheGlyThrValIleGlyArgSerArgAlaProGluVal	496
Db	1544	CGCACGCGAGTTTAGAGATTCTTTGGAAACAGAAATTCGTATCAAACTCCACAGTTGTA	1603
Qy	497	AlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMet	516
Db	1604	GCACGGTCTCTTCAAGGAGCCAAATTTTCTGCTTTGGAGATCTTGAACACGACTTG	1663
Qy	517	IleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeu	536
Db	1664	TTGGCTCCAGGAGTAAATATCTTTCGAGCTTGGACTGGAGACATGCACCATCGAGTTA	1723
Qy	537	ProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysPro	556
Db	1724	TCATCTGATCCAAAGGAGGTTAAGTCTCAATATATCTGTCGGAATCTTCAATGTCTAT	1783
Qy	557	HisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAla	576
Db	1784	CATGTAAGTGTGTAGCTGCTTTGATCAAAATCAAGGCATCCAGATTGAGTCTCGACGA	1843
Qy	577	IleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLys	596

Db	1844	ATCAAAATCAGCTCTCATGACACACATGCTTATGTTTCATGACACACATGTTTAAAGCCTCTTACG	1903
Qy	597	Asp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro	614
Db	1904	GATGCATCAGGAGCAGCTCTTCATCGCTTATGATCATCGGTGCAGGACATATAGATCCT	1963
Qy	615	GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr	634
Db	1964	TTAAGAGCTACAGATCCTGCTTTGGTCTACGACATTCGACCTCAAGAGTATTTTGAATTC	2023
Qy	635	LeuCystrHisLeuGlyPheThrArgSerAspIleLeuAlaIleThr---HisLysAsnVal	653
Db	2024	CTCTGCACTCAAGATTAAAGTCCATCACAGCTTAAAGTATTCAAAACATTCAAACAGA	2083
Qy	654	SerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAla	673
Db	2084	ACCTGCAACACACACTCTTGGCAAGAATCCGGGA---AACTTGAACCTACCCGGCGATATCA	2140
Qy	674	ValIlePheLysArgGlyLysThrThrGluMetIleThr-----ArgArgValThrAsn	693
Db	2141	GCTTTGTTCCAGAGAACACACATGTTTAAAGCTATGACACTTAGAAGACAGTGACCAAT	2200
Qy	692	ValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLysVal	711
Db	2201	GTTTGGTCTCACATTTCAAGCTACAGGTTCTGCTCTCGCCATCAAAAGGCGCATCCGTA	2260
Qy	712	IleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyr-----	729
Db	2261	ACTGTCAGGCCCAACACTCACTTCACCTCGNAGCACGAGAGCTTCTTACACGGTT	2320
Qy	730	-----ArgValTrpPheValLeuLysLysLysAsnArgGlyGlyLysValAlaSerPhe	747
Db	2321	ACTTTCAGGACAGGTTCCGGATCAAGAGGCTGAGTTTGGTGGT-----	2365
Qy	748	AlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerPro	767
Db	2366	-----CTAGTGTGGAAGAGCACT-----ACACATAAAGTTCGTAGCCCG	2404
Qy	768	IleSerValThr 771	
Db	2405	GTTATCATCA 2416	
RESULT 11			
ABZ13381			
ID	ABZ13381	standard; DNA; 2295 BP.	
XX	AC	ABZ13381;	
XX	DT	21-JAN-2003 (first entry)	
XX	DE	Arabidopsis thaliana stress regulated gene SEQ ID NO 1186.	
XX	KW	Arabidopsis thaliana; plant; gene; stress; transgenic; ds.	
XX	OS	Arabidopsis thaliana.	
XX	PN	WO200216655-A2.	
XX	PD	28-FEB-2002.	
XX	PF	24-AUG-2001; 2001WO-US26685.	
XX	PR	24-AUG-2000; 2000US-227866P.	
XX	PR	26-JAN-2001; 2001US-26464P.	
XX	PR	22-JUN-2001; 2001US-30011P.	
XX	PA	(SCRI) SCRIPPS RES INST.	
XX	PA	(SYGN) SYNGENTA PARTICIPATIONS AG.	
XX	PI	Harper JF, Kreps J, Wang X, Zhu T;	
XX	DR	WPI; 2002-304127/34.	
XX			

Identifying a stress condition to which a plant cell has been exposed
and producing plants with increased tolerance to these abiotic stresses

Claim 144; SEQ ID NO 1186; 577pp + Sequence Listing; English.

The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:
(a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and
(b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used in methods of the invention.

Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

[illegible]

```
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144633.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154013.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155136.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 1.72e-117 Length: 2489
Score: 1589.00 Matches: 342
Percent Similarity: 60.36% Conservative: 127
Best Local Similarity: 44.02% Mismatches: 268
Query Match: 39.55% Indels: 40
DB: 21 Gaps: 16

US-09-806-767-2 (1-775) x AAC42230 (1-2489)

QY 7 PheLeuCysIleIlePheLeuLeuPheCysSerSerSerGluIleuGlnLysGln 26
Db 225 TTCTCTCTCTTTCCGTTTATCTTTCCGAGCTTCTCAGGCC-----CCGAAG 272
QY 27 ThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPhe 46
Db 273 ACTTTCATTTCCGTTATCGATCGTGTGATCT---ATGCCTTCTATTATTTCCGCGCAGCTAC 329
QY 47 AspTyrHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGlu 66
Db 330 CATTTGGTAT-----AGCACCGAGTTCGCCGACGAA--- 359
QY 67 ProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaIleGlnLeu 86
Db 360 -----TCTCGAATCGTCCATGTTTACCACACAGCTTCCATGCTTCTCGCGCTTTGTT 413
QY 87 ThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArgProAsp 106
Db 414 ACTCCAGATGAACAGATAATCTCGTAACACCCAGCAGTCTCTGTGTTTTCGAAGAC 473
QY 107 HisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGly 126
Db 474 CGACGTCGAGAGCTTCACACCACAGTCTCTCTCAATTCTTGTGTTTACAA-----AAC 527
QY 127 AsnSerGlyValTyrSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAsp 146
Db 528 CAAAAGGACTATGTCAGATCTGATTACGGATCAGACGTAATCATTTGGGCTTTTTCAC 587
QY 147 ThrGlyValTyrProGluSerProSerPheAspThrGlyMetProSerIleProArg 166
Db 588 ACCGGAATTTGGCCGAGGAGGAGTTCAGATCTTACCTCGTCCAAATTCCTCAAAA 647
QY 167 LysTyrLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLys 186
Db 648 AGTGGAGAGGGGTTTGGGATCCGGAGCCAGATTCAGTCTCTCGGAATGTACCGCTAAA 707
```


PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145152.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149446.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.


```

Db 1506 GCTGTTGCAATATTTCTTCCAGGAGCCTTCAAGCCTCACAGAAGCATTCTCAAACT 1565
Qy 515 AspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsn-----Leu 531
Db 1566 GACATTACCGCAGCAGGAGTCTCGATACCTCGCTGCGATGGACTGGAAACGACTCAAGCATT 1625
Qy 532 GlyProThrGlyLeuProTyArgSerArgValAsnPheThrValMetSerGlyThr 551
Db 1626 TCACTGGAAGCAGCCGGTTCT-----CAGTATAACGTCATATCAGGAAT 1673
Qy 552 SerMetSerCysProHisValSerGlyIleThrAlaIleuIleArgSerAlaTrpProAsn 571
Db 1674 TCCATGCGAGCTCTCATGTTTTCAGCTGTTGCACTCTGATCAATCACAGCATCCACA 1733
Qy 572 TrpSerProAlaIleLysSerAlaLeuMetThrThrAla---AspLeuTyArgArg 590
Db 1734 TGGGTCATCCCGCATCAGATCAGCAATATGACACAGCAGCTCAAAACAACAGCAG 1793
Qy 591 GlnGlyLysAlaIleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGly 610
Db 1794 AAAGGCTTTATACAAAGAGAACTGGTGCAACAGCCACACTTATGACTCTGGAGCAGA 1853
Qy 611 HisValAsnProGlnLysAlaIleAsnProGlyLeuValTyAsnIleGlnProValAsp 630
Db 1854 GAATAGCTCAACAGCATCAATCAACAGGACTAGTTTACGAGACTACTGAACTGAC 1913
Qy 631 TyrIleThrTyLeuGlyThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHis 550
Db 1914 TACCTGAACCTTCTCTGTTACTATGATATACGTAACCAACAATAAAGGCTATGTCAAAA 1973
Qy 651 -----LysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeu 667
Db 1974 GCTTTTCCAGAGAATTTTACTTGCCCTGCAGATTCACACTTAGACTTGATCTCCACCATC 2033
Qy 668 AsnTyProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArg 687
Db 2034 AATTACCGCTCAATGGTAATCTCGGATTCAGAGAAATGGTAGCAGACAGTACAGA 2093
Qy 688 ArgValThrAsnValGlySerPro---AsnSerIleTySerValAsnValLysAlaPro 706
Db 2094 ACAGTGACCAATGTTGGTGAGGACGGTGAGCGGTTTACACAGTCAGCGTGAACACCA 2153
Qy 707 GluGlyLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThr 726
Db 2154 CCAGGGTTTAACTTCAAGTCAGCCAGCAAACTCAATTTACAAAAGATGGTGAGAAG 2213
Qy 727 LeuSerTyArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAlaSer 746
Db 2214 TTGACATACCAAGGTGATAGTG-----TCTGCTACTGCTTCA 2249
Qy 747 PheAlaGln-----GlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArg 763
Db 2250 CTTAAGCAAGATGATTTGGGGCTCTTACTTGGTCTAATGCC-----AAGTACAAG 2300
Qy 764 ValArgSerProIleSerValThrLeuLysThrAsn 775
Db 2301 GTCAGAACCCCAATGTAAATAGTAGCGAGAGTAGC 2336

RESULT 14
ID AAQ73756 standard; cDNA to mRNA; 2552 BP.
XX AC AAQ73756;
XX AC AAQ73756;
XX 21-JUL-1995 (first entry)
DE Gene encoding melon protease from Cucumis melo L.
KW Melon; protease; Cucumis; immunoscreen; antibody; pGEMEX-1; E.coli; ss.
OS Cucumis melo (L.) cv Earls Favourite.
XX

```

```

PH Key Location/Qualifiers
FT CDS 22..2217
FT /*tag= a
FT /product= melon protease
FT sig_peptide 22..351
FT /*tag= b
FT mat_peptide 352..2214
FT /*tag= c
PN JP06284890-A.
XX 11-OCT-1994.
XX 01-SEP-1992; 92JP-0233935.
XX 01-SEP-1992; 92JP-0233935.
XX (AJIN ) AJINOMOTO KK.
XX WPI; 1994-362597/45.
XX P-PSDB; AAR60822.
XX DNA encoding melon protease - used for recombinant production of
PT the enzyme, used in the food and chemical industries
PT Claim 4; Page 6-8; 9pp; Japanese.
XX
XX The nucleotide sequence of the gene encoding the novel melon protease
CC from Cucumis melo (L.) cultivar Earls Favourite. The gene sequence of
CC 2552 bp encodes a protein of 731 amino acids. The mature protein,
CC lacking a 110 a.a. signal peptide, has 621 a.a and mol. wt around 67 kD.
CC The gene for the protein was isolated from a melon cDNA library by
CC immunoscreening with an antibody. A partial clone was obtained which
CC was used, to rescreen the cDNA library, and isolate a clone containing
CC an almost full-length cDNA insert. The sequence of the insert was
CC determined. The full length DNA encoding the melon protease was isolated
CC from the DNA. The 2.6 kb coding fragment was subcloned into the plasmid
CC pGEMEX-1 for expression of the protein in E.coli when transformed into
CC E.coli JM109(DE3). The expressed protease can be use in the food and
CC chemical industries.
XX
SQ Sequence 2552 BP; 694 A; 533 C; 550 G; 775 T; 0 other;

Alignment Scores:
Pred. No.: 4.3e-84 Length: 2552
Score: 1172.50 Matches: 294
Percent Similarity: 51.52% Conservative: 114
Best Local Similarity: 37.12% Mismatches: 299
Query Match: 29.18% Indels: 85
DB: Gaps: 23

US-09-806-767-2 (1-775) x AAQ73756 (1-2552)
Qy 1 MetGluProLysProPhePheLeuLeuCysIleIlePheLeuLeuPheCysSerSer----- 18
Db 22 AATGCTTCTTCTCTAATCTTCAAGCTTTCTTTTAGCTCTTTCTTCTCAGTAATCGATC 81
Qy 19 -----SerSerGluIleLeuGlnLysGlnThrTyIleValGlnLeuHisPro 34
Db 82 GCTTCTAGATTGGATTCTGACGATGATGAAAAACATTATATTGTATACATGGGAGG 141
Qy 35 AsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeuSerPheLeuGln 54
Db 142 AAGTAGAGGATCTCTGAT-----TCTGCTCATTTTACATCATAGGCAATGTGGAA 192
Qy 55 GluAlaValLeuGlyValGluGluGluGluGluProSerSerArgLeuLeuTySer 74
Db 193 CAAGTTGTGGCAGCAGCTTTTCTCTCCAGAACT-----GTGCTCCACACT 237
Qy 75 TyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluLeu 94
Db 238 TACAAGAGAAGTTTCAACGGATTCCGAGTGAACACTTACTGAAGAAGAAGCTGAAAGATT 297

```

Qy 95 ArgTyrSerProGluValValAlaValArgProAspHisValLeuGlnValGlnThrThr 114
Db 298 GCCAGTATGAGGGTGTGGTCTGTCTGTTTAAATGAATGAACAACTTCATACGACA 357
Qy 115 TySerTyrLysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSer 134
Db 358 AGATCATGGGATTTCTGGGTTTTCCTCA-----CTAACTGTCTCTCGTCGAAGT 405
Qy 135 ArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTrpProGluSerPro 154
Db 406 CAAGTGAAGCAACATAGTTGTGGAGTTTGGACACCGAATCTGGCCGAATCTCCC 465
Qy 155 SerPheAspThrGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGlu 174
Db 466 AGTTTCACGATGAAGGTTTCAGTCTCCACCAACCAATGGAAGGCACTTGTGAACC 525
Qy 175 GlyGluSerPheSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIle 194
Db 526 TCCAACAACTTT-----CGTTGCAACAGAAAAATTATTGGAGCTGCATCATAC 576
Qy 195 ArgGlyHisArgValAlaAsnSerProGluSerProAsnMetProArgGluTyrIle 214
Db 577 ATAGGCGT-----CCCATTCACCCGGTGAATGTGAAT 609
Qy 215 SerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyLysSerSer 234
Db 610 GGTCACAGACACAAATGACACGCGGACGACACACTGCATCAACAGCGGCTGGTGTCTA 669
Qy 235 ValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGly 254
Db 670 GTTAGCCAGCAAACTCTATACGGTCTCGGCTCGGACGCGCAAGAGAGAGTTCCTCTTA 729
Qy 255 AlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeu 274
Db 730 GCGGAATCCTGCATACAGGATGCTGGAAATGATGGTGTCTCTGCATACAGACATCTT 789
Qy 275 AlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGly 294
Db 790 GCAGCATATGACGATGCCATTCGGGATGGAGTGGATCGATATATATCTCTTCAGTGGGTGG 849
Qy 295 Phe-----ProIleProLeuTyrAspThrIleAlaIleGlyThrPheArgAlaMetGlu 313
Db 850 GCTAAATCCAGCAGCATATTTCTGGTATGTCCTATGCCATGGATCTTTCCATGCGAGTAG 909
Qy 314 ArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAla 333
Db 910 AGAGGAATATTAACATCCAAATCTCTGCGGAATGGAGGCCCTAAATTTCTTCACCCGCA 969
Qy 334 AsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgPheProAla 353
Db 970 AGCTGTCTCCGGCTCTGTCTGTGTGCGAAGCACCATGGACAGAAAGTTTGTACA 1029
Qy 354 ValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyLysSerLeu----- 368
Db 1030 CAAGTACAGATTGGTAATGACAGAGCTTTCAGGAGTTCAATTAACATTTGATAAT 1089
Qy 369 -----TyrPro-----GlyLysGlyIleLysAsnAlaGlyArgGluValGluVal 383
Db 1090 CAATACATATCCCTTGTAGTGGCGGTATATACCAAT----- 1128
Qy 384 IleTyrValThrGlyAspLysGly--SerGluPheCysLeuArgGlySerLeuPro 402
Db 1129 -----ACTGGTTTCGATAAGTCCACCTCAAGGTTCTGACCGAACAAGTCAGTGAAT 1179
Qy 403 ArgGluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGlu 422
Db 1180 CCCAAATTGTAAGGGAATAATTTGTTGTGTAACCGAGTTTC-----GGTCTCATGAA 1236
Qy 423 LysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIle 442
Db 1237 TTCTTTAAGTCTTGGATGGTGCAGCGGGTGTCTCATG---ACATCAAAATACG----- 1287
Qy 443 AsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThr 462

RESULT 15

AAZ29905

ID AAZ29905 standard; cDNA; 2448 BP.

XX

AC AAZ29905;

XX

Db 1288 -----AGGGAATTATGCAGACTCCTATCCCTTGCCTTCTTCCGTCTCGACCCCAAT 1338
Qy 463 GluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIle 482
Db 1339 GATCTCCTTGCCTTTCGCTTATATTATTCATTCGCTCTCTCGGTGCA---ACCAT 1395
Qy 483 PheGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArg 502
Db 1396 TTCAAGAGTACCAAACTCTCAATGCTCTGCACTGTGTGTGTCTCTCTCTCATCAGG 1455
Qy 503 GlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsn 522
Db 1456 GGTCTAATCTGTGAACCTAAAGATGTTATTAAAGCCAGACATAAGTGGTCCAGAGTCCAA 1515
Qy 523 IleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArg 542
Db 1516 ATTCTAGCAGCATGGCCT---TCTGTTGCACCATTTGGT-----GGAATCCGTAGAAAC 1566
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1567 ACACTTTATATAATCTCAGGAACATCAATGCTTGTCCACATATCCTCGGAATTGCA 1626
Qy 563 AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaIleLysSerAlaLeuMet 582
Db 1627 ACCTACGTTAAACATACATACTACTTGTGTCTCTCTGCTCCATCAAGTCAGCACTCATG 1686
Qy 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1687 ACAACCGCTTCACCCATGAATGCTAGG-----TTCAATCCACAG 1725
Qy 603 GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeu 622
Db 1726 CGAGATTTCATATGTTTCAGGCCATGTTAACCCGTTAAAGCAGTAAGACCTCGGTTG 1785
Qy 623 ValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArg 642
Db 1786 GTTATGATGCAAAATGAAGCGACTACGTTAAATCTTGTGTGTGCTCAAGTTACACACC 1845
Qy 643 SerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsn 662
Db 1846 CAGCGGTTTCAGCTATCACCAGGACTATAGTCTTGTACTTCT-----GGTAAT 1896
Qy 663 ProGly-----PheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 1897 ACTGGTAGATGGGATTTAAACTATCTCTTTTGGACTTTCAGTATCTCTTCACAG 1956
Qy 681 Thr---ThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyr 699
Db 1957 ACTTTCATCACTACTTCACAGAACTCTCAGAGTGTCCGCCCTCAAGCATCAACATAT 2016
Qy 700 SerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuVal 719
Db 2017 AGAGCTATGATCTCTGCCACAAAGGCTTACTCTCAGTGAATCTTAATGTTCTATCA 2076
Qy 720 PheLysHisValAspGlnThrLeuSerTyrArgValTrpPheValLeuLysLysAsn 739
Db 2077 TTTAATGCGCTTGGAGATAGAAAATCTTTTACCTTGACAGTT----- 2118
Qy 740 ArgGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsn 759
Db 2119 CGAGATCAATAAAGGATTTGTAGTGTCTAGCTTCTTTGGTGTGGAGGATGGT----- 2172
Qy 760 LeuMetGlnArgValArgSerProIleSerValThr 771
Db 2173 ---GTACACTATGTGAGAGCCCTATAACCATCACA 2205
RESULT 15
AAZ29905
ID AAZ29905 standard; cDNA; 2448 BP.
XX
AC AAZ29905;
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 02:06:41 ; Search time 101 Seconds

(without alignments)
3386.851 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEKPFFLCIIIFLLFCSSS.....NSHNLQVRSPISVTLKTN 775

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09806767/runat_02022004_085959_17923/app_query.fasta_1.967
-DB=Issued Patents NA -CPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=100 -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09806767 @Cgn1_1_44 @runat_02022004_085959_17929 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	430	10.7	2532	1	US-07-671-376C-4
2	345	8.6	1977	3	US-08-894-818B-6
3	332	8.3	1962	3	US-08-894-818B-34
4	332	8.3	1962	4	US-09-445-472-15
5	330	8.2	1977	3	US-08-894-818B-2
6	330	8.2	1977	4	US-09-445-472-11
7	324.5	8.1	1236	4	US-09-445-472-2
8	324.5	8.1	1566	3	US-08-894-818B-4
9	297	7.4	2539	3	US-09-000-016-3
10	297	7.4	2539	4	US-09-514-340-3
11	297	7.4	2809	3	US-09-000-016-1
12	297	7.4	2809	4	US-09-514-340-1

13	271.5	6.8	1494	3	US-09-255-502-1	Sequence 1, Appli
14	271.5	6.8	1497	6	5472855-1	Patent No. 5472855
15	271.5	6.8	1868	1	US-08-069-863-1	Sequence 1, Appli
16	271.5	6.8	1868	1	US-08-309-069-1	Sequence 1, Appli
17	271.5	6.8	1868	1	US-08-562-833-1	Sequence 1, Appli
18	271.5	6.8	1868	4	US-09-659-749-1	Sequence 1, Appli
19	271.5	6.8	1868	4	US-09-672-105-1	Sequence 1, Appli
20	271.5	6.8	1868	5	PCT-US95-05520-1	Sequence 1, Appli
21	265.5	6.6	1146	2	US-08-504-265B-74	Sequence 74, Appli
22	264.5	6.6	8119	1	US-08-460-343B-1	Sequence 1, Appli
23	264.5	6.6	8119	1	US-08-398-028B-1	Sequence 1, Appli
24	264.5	6.6	8119	2	US-08-504-265B-1	Sequence 4, Appli
25	263.5	6.6	1140	3	US-09-023-173-4	Sequence 5, Appli
26	261	6.5	1330	4	US-09-966-921A-5	Sequence 6, Appli
27	260.5	6.5	1497	1	US-08-322-677A-6	Sequence 6, Appli
28	260.5	6.5	1497	1	US-08-322-676-6	Sequence 6, Appli
29	260.5	6.5	1497	3	US-08-898-218-6	Sequence 6, Appli
30	260.5	6.5	1497	3	US-08-848-793-6	Sequence 6, Appli
31	260.5	6.5	1497	4	US-09-445-270-1	Sequence 1, Appli
32	260.5	6.5	1497	4	US-09-178-173A-1	Sequence 1, Appli
33	260.5	6.5	1497	4	US-08-322-678-6	Sequence 6, Appli
34	260.5	6.5	1513	4	US-09-178-155-1	Sequence 1, Appli
35	259	6.4	1306	4	US-09-966-921A-1	Sequence 1, Appli
36	258.5	6.4	1496	6	RE34606-1	Patent No. RE34,60
37	257	6.4	2760	4	US-09-486-192-1	Sequence 1, Appli
38	253	6.3	1500	1	US-08-322-965-1	Sequence 1, Appli
39	252.5	6.3	10216	2	US-08-875-154-1	Sequence 1, Appli
40	248	6.2	1140	1	US-08-322-677A-11	Sequence 11, Appli
41	248	6.2	1140	1	US-08-322-676-11	Sequence 11, Appli
42	248	6.2	1140	3	US-08-898-218-11	Sequence 11, Appli
43	248	6.2	1140	3	US-08-848-793-11	Sequence 11, Appli
44	248	6.2	1140	4	US-09-445-270-6	Sequence 6, Appli
45	248	6.2	1140	4	US-08-322-678-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1

US-07-671-376C-4
; Sequence 4, Application US/07671376C
; Patent No. 5294542
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan et al.
; TITLE OF INVENTION: RESIDUAL PROTEASE-III
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/671,376C
; FILING DATE: 19910319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00811/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
 LENGTH: 2532 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-07-671-376C-4

Alignment Scores:

Pred. No.: 4.51e-33 Length: 2532
 Score: 430.00 Matches: 198
 Percent Similarity: 39.13% Conservative: 135
 Best Local Similarity: 23.27% Mismatches: 302
 Query Match: 10.70% Indels: 216
 DB: 1 Gaps: 35

US-09-806-767-2 (1-775) x US-07-671-376C-4 (1-2532)

Qy 7 PheLeuCysIleIlePheLeuPheCysSerSerSerGluIleLeu----- 23
 Db 61 TTTCGCTTGAAGTTTCGCTTATTTTTCGCTTATCCACAGGCATTACGGCGCTCAG 120
 Qy 23 ----- 23
 Db 121 GCAGCTCGGCTTCTTCAAAAACGTCGGCTGATCTGGAAGAAAGCCGAGGTATTTCGGTGAT 180
 Qy 24 ----- 36
 Db 181 ATCGATATGACGACAAAGAAACAAACCGTTATAGTGAATTAAGAAATCCTTG 240
 Qy 37 -----GluThrAlaLysThr 41
 Db 241 CGAAGCGGAGGAGGAGGAGGAGCAATCGAAGAAAGCAAGCTGAAACCGCTCGACC 300
 Qy 42 PheAlaSerLysPheAspTrpHisLeuSerPheLeuGlnGluAlaValGlu 61
 Db 301 AAAGCAAAA-----AACAAAGCAATCAAGACAGTGAGAA 333
 Qy 62 GluGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGly 81
 Db 334 -----AACGGAAAGTAAACCGGGAATATGAGCAGGTATTTCAGGC 375
 Qy 82 PheAlaAlaGlnLeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValVal 101
 Db 376 TTCTCTATGAAGCTTCAGCTATGAGATTCACAAACTTTCAGCGTAAAGACGTTAAG 435
 Qy 102 AlaValArgPro-----AspHisValLeuGlnValGlnThrTyr 115
 Db 436 CGAGTGACCGAAGCTCACATATAAACACAGACATATGAGGATAAAGACGTCAACATC 495
 Qy 116 SerTyrLysPheLeuGly-----LeuAspGly-----PheGlyAsnSerGly 129
 Db 496 TCCGAAGACGCGGTATCTCCGCAATATGATGACAGTGGCGCTTATATCGGAGCAACGAT 555
 Qy 130 ValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal 149
 Db 556 GCATGGGATTTAGGCTACACAGGAAAGGCAATCAAGGTGGCGATATTGACACTGGGGTT 615
 Qy 150 TrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLys 169
 Db 616 GAATACAAATCAACAGATCTGAAG-----AAA 642
 Qy 170 GlyIleCysGln-----GluGlyGluSerPheSerSerSerCysAsnArgLysLeuIle 188
 Db 643 AACTTTGGACAATATAAGGATACGATTTTGGACAATGATTACGAT----- 690
 Qy 189 GlyAlaAlaArgPheIleArgGlyHisArgValAlaAsnSerProGluGluSerPro--- 207
 Db 691 -----CCAAAGAAAGAACACCAACC 708
 Qy 208 AsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAla 227
 Db 709 GCGCATCCGAGG-----GGCAGGCAACTGACCATGGGCACACACGTAGCC 753

Qy 228 SerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyVal 247
 Db 754 GGAACCTGTGGCTGCAAC-----GGAAACG 777
 Qy 248 AlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGly 267
 Db 778 ATTAAGCGCTAGCGCTGATGCCACACTTCTTCTTATCTGTTTATGAGCCCTGGCGGA 837
 Qy 268 CysTyrSerSerAsp---IleLeuAlaAlaIleAspValAlaIleGlnAspLysValAsp 286
 Db 838 AGCGGCACACGGAACAGTCATCGCGGGGTGGAACCTGCAGTGCAGGACGGGGCAGAT 897
 Qy 287 ValLeuSerLeuSerLeuGly---GlyPheProIleProLeuTyrAspThrIleAla 305
 Db 898 GTGATGAACCTGTCTCTCGGAACCTCTTAAACAACCCGGACTGGGCGACACACAGCG 957
 Qy 306 IleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsn 325
 Db 958 CTTGACTGG-----GCCATGTCAGAAAGCGTGTCTGCTGTACTCTCAACGCGCAACAGC 1011
 Qy 326 GlyProIleGluSerSerValAlaAsn-----ThrAlaProTrpValSerThrIleGly 343
 Db 1012 GGACCGAACGCTGCAGCTCGGATCGCGGCGCATCAAGAGAGCGATTCTTCTCGGT 1071
 Qy 344 AlaGlyThrLeuAspArgPheProAlaValVal-----ArgLeuAlaAsnGlyLys 361
 Db 1072 GCGACTAGCTGCGCTCAATAGTACCGCTCACTTTCGGCTCCTACTCTTCAGCAAAA 1131
 Qy 362 LeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluVal 381
 Db 1132 GTGATGGCTACAAACAAGAGGACGAGCTCAAGCGCTCAATAAC-----AAAGAAGTT 1185
 Qy 382 GluValIleTyrValThrGlyLysAspGlySerGluPheCysLeuArgGlySerLeu 401
 Db 1186 GAGCTTGTGTC---GAAGCGGGAATCGGCAAGCAAGGATTTT----- 1224
 Qy 402 ProArgGluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSer 421
 Db 1225 GAAGGGAAGACCTGACAGGCAAGTCCCGTGTCTCAACGAGGACGATTCATTGTGTG 1284
 Qy 422 GluLysGlyGluAlaValLysGluAlaGlyValAlaMetIleLeuAlaAsnThrGlu 441
 Db 1285 GATAAAGCGGATACGCTAAAGACCGGTGCAATCGSCATGGTGTGTATTAACACCTC 1344
 Qy 442 IleAsnGlnGluLysSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyr 461
 Db 1345 TCTGAGAAATGAAGCCATGTGCGGCGATGTCTGCCAACGATTAAGCTTTCATTA 1404
 Qy 462 ThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIle 481
 Db 1405 GAAGACGCGCAAAACTC-----GTCAGCGCTGAAAGCTGTGAGACAAAAACA 1455
 Qy 482 IlePheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAla 501
 Db 1456 ACATTCAAGTTGACCGTCT---TCAAAGCGCTCGGTGACCAAGCTCGCTGATTTCTATCA 1512
 Qy 502 ArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyVal 521
 Db 1513 CCGCGCCCT---GTTATGATACGTGGATGATTAAAGCTGATATTTCGCGCCAGGGGTC 1569
 Qy 522 AsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArg 541
 Db 1570 AATATCGTGAGCAGCATCCCAACACGATCCTGACCATCCATCGGCTACGGTACGATCAAA 1629
 Qy 542 ArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIle 561
 Db 1630 CAA-----GGAAACAGCATGGCATCGCTTCATATTTCGCGGAGCG 1668
 Qy 562 ThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeu 581
 Db 1669 GTTGGCTTATTAAACAGCCCAACCAAGTGGAGCGTTGAACAGATTAAAGCCCATC 1728
 Qy 582 MetThrThrAla---AspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLys 600

Db 1729 ATGAATACCGCTGTCACTTTAAAGGATAGCGATGGGGAAGATATATCGCATAAC-----1782
Qy 601 ProlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 1783 -----GTCAGGCGCGAGCGAGCAAGAAATTAAGACGCAATCAAGGCC 1827
Qy 621 GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGly---639
Db 1828 GATTGCTGCTCACTACCTGGAGCTATTATCATCGGCACGTTCTTTGAAGGAAACGGAAC 1887
Qy 640 PheThrArgSerAspIleLeuAlaIleThrHisLysAsn-----652
Db 1888 GAAACAAAATGAAGCGTTTACGATTCAAAATCAATCTCCATTAGAAAGTCATACACA 1947
Qy 653 -----ValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeu-----667
Db 1948 CTTGAATACTCATTTAATGGCAGCGCATTTCCACATCCGCGCACAGCCGTGTGTGATT 2007
Qy 667 -----667
Db 2008 CCGGCAATCAAAACGGGAAAGCCACTGCAAAAGTAAAGTCAATACGAAGAAACAAA 2067
Qy 668 -----AsnTyrProSerIleAlaValIlePheLysArgGlyLysThr-----681
Db 2068 GCTGGCACCATTAGAGNACGGTTATCGTCAGAGAGCGGAAACGGTCGCTAAGGTA 2127
Qy 682 ---ThrGluMetIleThrArg-----ArgValThrAsnVal-----692
Db 2128 CCTACATTGCTGATTGTGAAGAGCGCGATTATCCGAGAGTCACATCTGCTCTGTGACG 2187
Qy 693 ---GlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGly-----708
Db 2188 GAAGGGTCTGTACAAAGTACTATCAAAATTGAAACCTACCTTCTCGGAGCGGAAGAG 2247
Qy 709 IleLysValIleValAsnProLysArgLeu-----ValPheLys 721
Db 2248 CTGGCGTCTCTGCTCTATGACAGCAACCTTGATTTCGACGCCAAGCCGCAATTTATAA 2307
Qy 722 HisValAspGlnThrLeuSerTyrArgValTyr 732
Db 2308 AACCAAGATAAAGGTTACCAGTACTTTGACTGG 2340

RESULT 2

US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-6

Alignment Scores:

Pred. No.: 1.14e-24 Length: 1977
Score: 345.00 Matches: 145
Percent Similarity: 33.17% Conservative: 62
Best Local Similarity: 23.24% Mismatches: 145
Query Match: 8.59% Indels: 272
DB: 3 Gaps: 21

US-09-806-767-2 (1-775) x US-08-894-818B-6 (1-1977)

Qy 29 lleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrp 48
Db 148 ATTCAAAATTAAGTCTTAACGAGGAAATCAGACAGTAATTTGTA-----TTTGAAGAA 201
Qy 49 HisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSer 68
Db 202 CAT-----ACGGAAGAAAGAAATTGCA 222
Qy 69 SerArgLeuLeuTyrSerTyrGlySer-----AlaIleGluGly 81
Db 223 GTAAGAGTCTTGAGTTAATCGGTGCAAAAGTAGGTATGTACCATATTATATACCCGCA 282
Qy 82 PheAlaAlaGln-----LeuThrGluSerGlu 90
Db 283 ATAGCTGCCGATCTTAAGGTTAGACACTTACTAGTCATCTCAGGTTTAAACAGGGGTA 342
Qy 91 AlaGluIle-----LeuArgTyrSerProGluValValAlaValArg-----104
Db 343 GCTAAGCTTTTCCAGGTGTAGGTTTATCCAGGAAGACTACAAAGTTACAGTTTCAGCAGAA 402
Qy 105 -----ProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeu 120
Db 403 TTAGAAGACTGGATGAGTCTGCAGCTCAAGTTATGGCACTTAC-----447
Qy 121 GlyLeuAspGlyPheGlyAsnSerGlyValTyrSerLysSerArgPheGlyGlnGlyThr 140
Db 448 -----GTTTGGAACTTGGGATATGATGTTCTGGAATC 480
Qy 141 lleIleGlyValLeuAspThrGlyValTyrProGluSerProSerPheAspThrGly 160
Db 481 ACAATAGGAATAATTGACACTGGAATT-----GACGCTTCT 516
Qy 161 MetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSer 180
Db 517 CATCCAGATCTC-----528
Qy 181 SerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200
Db 529 -----CAAGGAAAGTAATTGGGTGGGTAGATTGTCATGTT-----567
Qy 201 AsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThr 220

Db 568 -----AGAGTTAT-----CCATACGATGACCAT 591
Qy 221 GlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaVal 240
Db 592 GGACATGGAACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCA-----639
Qy 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyr 260
Db 640 -----AGTAATGGCAAGTAGTCAAGAGGAATGGCTCCAGGAGCTAAGCTGGCGGAAT 690
Qy 261 LysValCysTrpPheAsnGlyCysTyrSer---SerAspIleLeuAlaAlaIleVal 279
Db 691 AAGCTTCTAGTCCGATGCTTCTGGAGCATATCTACTATAATTAATTAAGGAGTGGATGG 750
Qy 280 AlaIleGlnAspLys-----ValAspValLeuSerLeuSerLeuGlyGlyPhe 295
Db 751 GCCGTTGATACAAAGATAGTAGCGGAATTAAGGTCAATTAATCTTCTCTGTTCAAGC 810
Qy 296 ProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGly 315
Db 811 CAGAGCTCCGACGGAACCGACTCCCTCAGTCAGCGCGTCAACAAACCGCTGGAGCGCGT 870
Qy 316 IleSerValIleCysAlaIleGlyAsnAsnGlyProIleGluSerSerValAlaAsnThr 335
Db 871 ATAGTAGTCTGCGTCCGCGCGCAACAGCGCGCCGCAACACCTACACCGCTCGCTCACC 930
Qy 336 AlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaVal 355
Db 931 GCCGCCGCGAGCGAGTCAACCGTCGTCGTCAGTTGACAGCAAC-----975
Qy 356 ArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLys 375
Db 975 -----975
Qy 376 AsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPhe 395
Db 975 -----975
Qy 396 CysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArg 415
Db 975 -----975
Qy 436 IleLeuAlaAsnThrGluIleAsnGlnGluGluAspSerIleAspValHisLeuLeuPro 455
Db 975 -----975
Qy 456 AlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrVal 475
Db 975 -----975
Qy 476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db 976 -----GACAC 981
Qy 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
Db 982 ATGCCAGCTTCTCCAGCAGGGGACCGCGC-----GACGGAAGGCTCAAGCGGAA 1035
Qy 516 MetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThr---534
Db 1036 GTCGTCGCGCCCGCGCTTGACATCATAGCC-----CCGCGCGCCAGCGGAACCATG 1089
Qy 535 GlyLeuProTyrAspSerArgValAlaAsnPheThrValMetSerGlyThrSerMetSer 554
Db 1090 GGCACCCCGATACAGAC-----TACTACCAAGCGCTCTGGAACCATGAGCC 1140
Qy 555 CysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerPro 574

Db 1141 ACCCGCACGTTTGGCGGTTGGCGGCTCATCTCCAGGCCACCGAGCTGACCCCG 1200
Qy 575 AlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAla 594
Db 1201 GACAGGTGAAGACCGCCCTCATCGAGACCGCCGACATATGTC-----GCCCCAGGAG 1254
Qy 595 IleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db 1255 ATACCGGAC-----ATCGCTACGTCGGGTAGGTTGAACGTC 1293
Qy 615 GlnLysAlaIle 618
Db 1294 TACAAGGCCATC 1305
RESULT 3
US-08-894-818B-34
; Sequence 34, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikumoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1962 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-894-818B-34
Alignment Scores:
Pred. No.: 2,32e-23 Length: 1962
Score: 332.00 Matches: 140
Percent Similarity: 22.75% Conservative: 65
Best Local Similarity: 22.36% Mismatches: 147
Query Match: 8.26% Indels: 274
DB: 3 Gaps: 21

US-09-806-767-2 (1-775) x US-08-894-818B-34 (1-1962)

QY 29 IleValGlnLeuHisProAsnSerGluThrAlaIleValThrPheAlaSerLysPheAspTrp 48
Db 148 ATTCAGAAATGAAATCTTACGAGAAATCAGACACAGTAATTGTA-----TTTGAAC 201
QY 49 HisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluGluProSer 68
Db 202 CAT-----ACGGAAAGAAATGCA 222
QY 69 SerArgLeuLeuTySerTyGlySer-----AlaIleGluGly 81
Db 223 GTAAGAGTCTTGAGTTAATGGTGCAAAAGTTAGGTATGTACCATATATTATACCGCA 282
QY 82 PheAlaAlaGln-----LeuThrGluSerGlu 90
Db 283 ATAGTCCCGATCTTAAGTTAGAGCTTTACTAGTCATCTCAGGTTTAAACAGGGGTAA 342
QY 91 AlaGluIle-----LeuArgTySerProGluValValAlaValArg----- 104
Db 343 GCTAAGCTTTCAGTGTAGTTATCCAGGAAGCACTACAAGTTACAGTTTCAGCAGAA 402
QY 105 -----ProAspHisValLeuGlnValGlnThrThrTySerTyLysPheLeu 120
Db 403 TTAGAGGACTGGATGAGTCTGCAGCTCAAGTTATGCACTTAC----- 447
QY 121 GlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThr 140
Db 448 -----GTTTGAACCTGGGATATGATGTTCTGGAATC 480
QY 141 IleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGly 160
Db 481 ACAATAGGAATAATGACACTGGAATT-----GACGTTCT 516
QY 161 MetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSer 180
Db 517 CATCAGATCTC----- 528
QY 181 SerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200
Db 529 -----CAAGGAAAGTAATTTGGTGGTAGATTTTGTCAATGTT----- 567
QY 201 AsnSerProGluGluSerProAsnMetProArgGluTyLysSerAlaArgAspSerThr 220
Db 568 -----AGGAGTTAT-----CCATACGATGACCAT 591
QY 221 GlyHisGlyThrHisThrAlaSerThrValGlyGlySerValSerMetAlaAsnVal 240
Db 592 GGACATGGAATCATATGAGTCTCAATAGCAGCTGTTACTGGAGCAGCA----- 639
QY 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTy 260
Db 640 -----AGTAATGGCAAGTCAAGGAATGCTCCAGGAGTCAAGTGGCGGNAAT 690
QY 261 LysValCysTrpPheAsnGlyCysTySer-----SerAspIleLeuAlaIleAspVal 279
Db 691 AAGGTTTCAGGTGCGAGTGTCTGGAAGCATATCTACTATAATTAAGGAGTTGAGTGG 750
QY 280 AlaIleGlnAspLys-----ValAspValLeuSerLeuSerLeuGlyGlyPhe 295
Db 751 GCCGTGTGATAACAAAGATAAGTACGGAATTAAGTTCATTATCTCTCTGTTCAAGC 810
QY 296 ProIleProLeuTyArgAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGly 315
Db 811 CAGAGCTCAGATGTTACTGACGCTCTAAGTCAGGCTGTTAATGAGCGTGGATGCTGGA 870
QY 316 IleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThr 335
Db 871 TTACTGTGTGGTGGCTGGAACAGTGGACTAACAAGTATACATCGGTCTTCCA 930
QY 336 AlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValVal 355

Db 931 GCAGCTGCAAGCAAAAGTTATTACAGTTGGAGCCCTTGACAAG----- 972
QY 356 ArgLeuAlaAsnGlyLysLeuLeuTyGlyGluSerLeuTyProGlyLysGlyIleLys 375
Db 972 ----- 972
QY 376 AsnAlaGlyA-gGluValGluValIleTyValThrGlyGlyAspLysGlySerGluPhe 395
Db 972 ----- 972
QY 396 CysLeuArgGlySerLeuProArgGluIleLeuArgGlyLysMetValIleCysAspArg 415
Db 972 ----- 972
QY 416 GlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMet 435
Db 972 ----- 972
QY 436 IleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuPro 455
Db 973 -----TATGATGTT----- 981
QY 456 AlaThrLeuIleGlyTyThrGluSerValLeuLeuLysAlaTyValAsnAlaThrVal 475
Db 981 ----- 981
QY 476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db 981 ----- 981
QY 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
Db 982 ATACCAAGCTTCTCAAGCAGAGGCGCAACTGCA-----GACGGCAGGCTTAAGCCTGAG 1035
QY 516 MetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGly 535
Db 1036 GTTGTGTTCTCCAGGAACCTGGATATTCTGCCAGACGCAAGTGGAACTAGC---ATGGGT 1092
QY 536 LeuProTyArgSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCys 555
Db 1093 CAACCAATTAATGAC-----TATTACACAGCAGCTCTCGGACATCAATGGCAACT 1143
QY 556 ProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyProAsnTrpSerProAla 575
Db 1144 CTCACGAGTGGTATTGCGACCTCTGTCCAGCACACCCGAGCTGAGCTCCAGAC 1203
QY 576 AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyArgArgGlnGlyLysAlaIle 595
Db 1204 AAGTAATAAACACAGCCTCATAGAACTGCTGATATCGTA----- 1242
QY 596 LysAspGlyAsnLysProAlaGlyVal-----PheAlaIleGlyAlaGlyHisValAsn 613
Db 1243 -----AAGCCAGATGAAATAGCCGATATAGCTACGCTGACGTTAGGTGTTAAT 1290
QY 614 ProGlnLysAlaIleAsn 619
Db 1291 GCATCAAGGCTATAAAC 1308

RESULT 4

US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997

Db 1291 GCATACAGGCTATAAC 1308
 RESULT 5
 US-08-894-818B-2
 ; Sequence 2, Application US/08894818B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894,818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 323285/1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA=1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1977 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 US-08-894-818B-2
 Alignment Scores:
 Pred. No.: 3,74e-23 Length: 1977
 Score: 330.00 Matches: 124
 Percent Similarity: 34.54% Conservative: 57
 Best Local Similarity: 23.66% Mismatches: 119
 Query Match: 8.21% Indels: 224
 DB: 3 Gaps: 17
 US-09-806-767-2 (1-775) x US-08-894-818B-2 (1-1977)
 QY 103 ValArgProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeu 122
 Db 382 ATACGAGGATTACAAGGTTACAGTTCAGACGCCACTTCGGTCTCCCGATAGGGGCC 441
 QY 123 AspGlyPheGlyAsnSerGlyValTyrSerLysSerArgPhe---GlyGlnGlyThrIle 141
 Db 442 GAT-----ACCGTCTGGAACCTCCCTCGCTACGACGGAACGGGTGGTG 486

QY 142 IleGlyValLeuAspThrGlyValTyrProGluSerProSerPheAspThrGlyMet 161
 Db 487 GTTGCATCGTATCGGTTATA-----GACGGGAACCAAC 522
 QY 162 ProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSer 181
 Db 523 CCCGATCTG-----AAGGCG----- 537
 QY 182 SerCysAsnArgLysLeuIleGlyAlaArgPheIleArgGlyHisArgValAlaAsn 201
 Db 538 -----AAGTTCATAGGCTGTACGACGCGGTCAACGCGCAGGTGCG----- 576
 QY 202 SerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGly 221
 Db 577 -----ACCCCTACGATGACACGAGGA 597
 QY 222 HisGlyThrHisThrAlaSerThrValGlyGlySer---SerValSerMetAlaAsnVal 240
 Db 598 CACGGAACCCACGTTGCGGGTATCTTCCCGAACCAGCGAGGTTAACTCCAGTACATA 657
 QY 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyr 260
 Db 658 -----GGCGTGGCCCCGGCGGCAAGCTCGTCGGCGTC 690
 QY 261 LysValCysTrpPheAsnGlyCysTyrSer---SerAspIleLeuAlaAlaIleAspVal 279
 Db 691 AAGGTTCTCGGTGCGGACGTTGCGGAAGCTCTCCACCATCATCGCGGTGTGACTG 750
 QY 280 AlaIleGlnAspLys-----ValAspValLeuSerLeuSerLeuGlyGlyPhe 295
 Db 751 GTCGTCCAGAACAGGACAGTACGGGATAAGGTCATCAACCTCTCCCTCGGCTCCTCC 810
 QY 296 ProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGly 315
 Db 811 CAGAGCTCGACGGAACCGACTCCCTCAGTCAGGCGGTCAACACCGCTGGGACGCCGT 870
 QY 316 IleSerValIleCysAlaAlaGlyAsnGlyProIleGluSerSerValAlaAsnThr 335
 Db 871 ATAGTAGTCTCGCTGCGCGCGGCAACAGCGGGCGGAACACCTACACCGTGGCTCACCC 930
 QY 336 AlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValVal 355
 Db 931 GCCCGCGGACGACAGGTATACCGTCTCGTGCAGTGTACAGCAAC----- 975
 QY 356 ArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLys 375
 Db 975 ----- 975
 QY 376 AsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPhe 395
 Db 975 ----- 975
 QY 396 CysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArg 415
 Db 975 ----- 975
 QY 416 GlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMet 435
 Db 975 ----- 975
 QY 436 IleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuPro 455
 Db 975 ----- 975
 QY 456 AlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrVal 475
 Db 975 ----- 975
 QY 476 LysProLysAlaArgIleIlePheGlyThrValIleGlyArgSerArgAlaProGlu 495
 Db 976 -----GACAAC 981
 QY 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515

Db	667	GCTGCAGAGCAAGTGGAACTAGC---ATGGGTCAACCAATTAATGAC-----TAT	714
Qy	545	PheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeu	564
Db	715	TACACAGCAGCTCCTGGGACATCAATGGCACTCTCAGTAGCTGGTATTGCAGCCCTC	774
Qy	565	IleArgSerAlaTyProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThr	584
Db	775	TTGTCTCAAGCACACCCAGCTGGACTCCAGACAAAGTAAACACAGCCCTCATAGAACT	834
Qy	585	AlaAspLeuTyAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAlaGlyVal	604
Db	835	GCTGATATCGTA-----AAGCCAGATGAATA	861
Qy	605	-----PheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn	619
Db	862	GCGCATATAGCTACGTCAGGTAGGTTAATGCATACAGGCTATTAAC	912

RESULT 8

US-08-894-818B-4

Sequence 4, Application US/08894818B

Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: YAMAMOTO, Katsuhiko

APPLICANT: MITTA, Masanori

APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNASAWA, Susumu

APPLICANT: KATO, Ikunoshin

TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300

CITY: Washington

STATE: D.C.

COUNTRY: United States of America

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/894,818B

FILING DATE: 20-MAY-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03253

FILING DATE: 07-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 323285/1995

FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TAKAKURA-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1566 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

FEATURE:

OTHER INFORMATION: /note= N at position 1283 is G or T.

US-08-894-818B-4

Alignment Scores:	8.98e-23	Length:	1566
Pred. No.:	324.50	Matches:	119
Score:	32.88%	Conservative:	51
Best Local Similarity:	23.02%	Mismatches:	116
Query Match:	8.08%	Indels:	231
DB:	3	Gaps:	15
US-09-806-767-2 (1-775) x US-08-894-818B-4 (1-1566)			
Qy	110	GlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGlyAsnSerGly	129
Db	34	CAAGTTATGCAACTTAC-----	51
Qy	130	ValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal	149
Db	52	GTTTGGAACCTGGGATATGATGTTCTGGAATCACAATAGGAATAATGACACTGGAAT	111
Qy	150	TrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLys	169
Db	112	-----GACGCTTCTCATCCAGATCTC-----	132
Qy	170	GlyIleCysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGly	189
Db	133	-----CAAGGAAAAGTAATTTGGG	150
Qy	190	AlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMet	209
Db	151	TGGGTAGATTTTCTCAATGGT-----	171
Qy	210	ProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThr	229
Db	172	---AGGAGTTAT-----CCATAGCATGACCATGGACATGGAACCTCATGTAGTTCAATA	222
Qy	230	ValGlyGlySerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArg	249
Db	223	GCACCTGGTACTGGAGCAGCA-----AGTAATGCGAAAGTACAAG	261
Qy	250	GlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyr	269
Db	262	GGAATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCGGATGTTCTGGA	321
Qy	270	Ser---SerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLys-----	284
Db	322	AGCATATCTACTAATAATTAAGGGAGTTGAGTGGGCCCTTGATACAAAGATAAGTACCGA	381
Qy	285	ValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyAspAspThrIle	304
Db	382	ATTAAGTTCATTAATCTTCTCTGGTTCAGCCAGAGCTCAGATGCTACTGACGCTCTA	441
Qy	305	AlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsn	324
Db	442	AGTCAGGCTGTTAATGCAGCGTGGGATGCTGGATTAGTTGTTGGTTGCGCTGGAAC	501
Qy	325	AsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyVala	344
Db	502	AGTGACCTAAACAAGTATACAACTCGTTCTCCACAGCTGCCAACAAAGTATTACAGTT	561
Qy	345	GlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTy	364
Db	562	GGAGCGCTTGACAAG-----	576
Qy	365	GlyGluSerLeuTyProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIle	384
Db	576	-----	576
Qy	385	TyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGlu	404
Db	576	-----	576
Qy	405	GluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGly	424
Db	576	-----	576


```

QY 425 GluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln 444
Db 576 -----
QY 445 GluGluAspSerIleAspValHisLeuLeuProAlaThrIleuLeuGlyTyrThrGluSer 464
Db 577 -----TATGATGTT-----
QY 465 ValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGly 484
Db 585 -----
QY 485 GlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyPro 504
Db 586 -----ATAACAAGCTTCTCAAGCAGAGGCCCA 612
QY 505 SerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAlaIleIle 524
Db 613 ACTGCA-----GACGGCAGGCTTAAGCTGAGGTTGTTGCTCCAGGAAACTGGATAATT 666
QY 525 AlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsn 544
Db 667 GCTCCACAGCAAGTGAACTAGC---ATGGGTCAACCAATTATGAC-----TAT 714
QY 545 PheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeu 564
Db 715 TACACAGCAGCTCTCTGGGACATCAATGGCACTCTCAGCTAGTGTGTATTCAGCCCTC 774
QY 565 IleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThr 584
Db 775 TTGCTCCAAGCACACCCGAGCTGACTCCAGACAAAGTAAACACAGCCCTCATAGAAACT 834
QY 585 AlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAlaGlyVal 604
Db 835 GCTGATATCGTA-----AAGCCAGATGAATA 861
QY 605 -----PheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
Db 862 GCCGATATAGCTACGTCAGGTGACGGTAGGTTAATCATACAGGCTATAAAC 912

RESULT 9
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 203 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367

```

```

; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:

```

```

; INFORMATION FOR SEQ ID NO: 3:

```

```

; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 2539 base pairs

```

```

; TYPE: nucleic acid

```

```

; STRANDEDNESS: double

```

```

; TOPOLOGY: linear

```

```

; MOLECULE TYPE: genomic DNA

```

```

; ORIGINAL SOURCE:

```

```

; ORGANISM: Streptomyces viridosporus

```

```

; STRAIN: A-914

```

```

; FEATURE:

```

```

; NAME/KEY: CDS

```

```

; LOCATION: 338...2539

```

```

; IDENTIFICATION METHOD: E

```

```

US-09-000-016-3

```

```

Alignment Scores:

```

```

Pred. No.: 1,248-19 Length: 2539

```

```

Score: 297.00 Matches: 155

```

```

Percent Similarity: 31.03% Conservative: 70

```

```

Best Local Similarity: 21.38% Mismatches: 224

```

```

Query Match: 7.39% Indels: 276

```

```

DB: 3 Gaps: 22

```

```

US-09-806-767-2 (1-775) x US-09-000-016-3 (1-2539)

```

```

QY 36 SerGluThrAlaLysThrPheAlaSer---LysPheAspTrpHisLeuSerPheLeuGln 54
Db 632 GGGGACGCGCGCGCTGCTGGAGCGCGAGCTCGACCGCGGCTCTTCGACATCACC 691
QY 55 GluAlaValLeuGlyValIleGluGluGluProSerSerArgLeuTyrSer 74
Db 692 GAATCGCGCAGCGCGCAGCGCACTCCAGAAACAGGGACTGAAGTTCATCTCGCGC 751
QY 75 TyrGlySerAlaIleGluGlyPheAlaAlaGln----- 85
Db 752 TACCAGGGCGCGCAGCGCGCGCCAGCGCGAGGTCCCGAAGCGGGCGAACTCCGCGCG 811
QY 86 ---LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArg 104
Db 812 ACCCTGACGTCCCTGAACGCGGACCGGTGGGACCCCGCAGCAGGACGCGTCCGAGCTG 871
QY 105 ProAspHisValLeuGlnValGlnThrThrTyrSer---TyrLysPheLeuGlyLeuAsp 123
Db 872 TGGGACGCGCTCACCAACGCGCAGCGGACCGGACCGCCCTCGGCATCGCCACGCTCGGTGGAC 931
QY 124 Gly-----PheGlyAsnSerGlyValTrp 131
Db 932 GGGGTCCGCGCAGCGCGCGCTCGACACGTCTCGGGGACAGATCGGCGCCCAAGGCGTGG 991
QY 132 SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTrpPro 151
Db 992 TCCGCGCGCTACGACGCGCAAGCGCGTGAAGATCGCGCTCGGACACCGGTGTC----- 1045
QY 152 GluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyIle 171
Db 1046 -----GACACGAGCATCGGACCTG-----AAGGCGCGG 1075
QY 172 CysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLysLeuIleGlyAlaArg 191
Db 1076 GTGACCGCGCTCCAGAACTTCACCGCGCGCGCC----- 1108
QY 192 PhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArg 211
Db 1108 ----- 1108
QY 212 GluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGly 231

```

Db 1109 -----GGCGCCGGCGACAGGTGGCGCCACCGCACCGTCGCTCGATCGCGCG 1159
Qy 232 GlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMet 251
Db 1160 GGACAGCGGC-----GCCAGTCCAAGGCGCAAGTACAAGGCGCTC 1198
Qy 252 AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys--TyrSer 270
Db 1199 GCACCGCGCCCGGATCTCAACGGCAGAGTCTCGACGATCCGGTTTCGGCGACGAC 1258
Qy 271 SerAspIleLeuAlaIleAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeu 290
Db 1259 TCCGGCATCTCCCGCGCATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1318
Qy 291 SerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArg 310
Db 1319 AGCTCGCGCGCATGACACACCGGAGACCGACCGCTCGAGCGCGCGCTCGACAGGTG 1378
Qy 311 AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSer 330
Db 1379 TCCGCCGAGAGGCGGTCTGTTCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1438
Qy 331 SerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArg 350
Db 1439 GGTTCGCCCGCGAGCGCGCGCGCGCTCACGTCGCGCGC----- 1480
Qy 351 PheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrPro 370
Db 1480 ----- 1480
Qy 371 GlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAsp 390
Db 1481 -----GTCCAGCACAGGAC 1495
Qy 391 LysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMet 410
Db 1496 AAG----- 1498
Qy 411 ValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAla 430
Db 1498 ----- 1498
Qy 431 GlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAsp 450
Db 1498 ----- 1498
Qy 471 ValAsnAlaThrValLysProLysAlaArgIlePheGlyThrValIleGlyArg 490
Db 1498 ----- 1498
Qy 491 SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer 510
Db 1499 -----CTCGCGCGACTTCTCTCCACCGCGCGCGCGCTCGCGCGCGCGCGCG 1543
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaIleAlaIleProGln--- 529
Db 1544 ATC---AAGCGGAGCGTCACCGCTCCCGGCGTGACATCACCGCGCTCGCGCGAGGCG 1600
Qy 530 -----AsnLeuGlyProThrGlyLeuProTyrAspSerArgArg 542
Db 1601 AACGACATCGCGCCAGGAGTGGTGAGGACCGCGCGC----- 1639
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1640 -----TACATGACCATCTCCGCGACCTCGATGGCGACCGCGCGCGCGCGCGCGCG 1693
Qy 563 AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaIleLysSerAlaLeuMet 582
Db 1694 GCCCTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1750

Qy 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1751 -----ACCGCTCCACCAAGGC 1768
Qy 603 Gly-----ValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
Db 1769 GGCAAGTACACCCCGTTTCGAGCAGGTTTCGGCGCGATCCAGGCCGACAGGGCGCTCCAG 1828
Qy 620 ProGlyLeuVal-----TyrAsnIleGln----- 627
Db 1829 CAGACCGTGTATCGCGCACCGCGTCTCGGTGAGTTCGGCTCCAGCAGTGGCGCGCAC 1888
Qy 628 -----ProVal---AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSer 643
Db 1889 GACGACGAGCGGTTCACCAAGCAGCTGACCTACGCG---AACCTCGCG---ACCAGGAC 1942
Qy 644 AspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnPro 663
Db 1943 GTACAGCTGAAGCTGACGTGCGACCGCCACCGCACCGCAAGCGCGCGCGCGCGCGCG 2002
Qy 664 GlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrGlu 683
Db 2003 TTCTTCAGCTGGCGCGCACCGCTGACCGTCCCGCGCGCGCGCGCGCGCGCGCGCG 2062
Qy 684 MetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnVal 703
Db 2063 ATG---ACCGCGCACCGCGCTCGCGCGCACCGGTGACCGCGCGCTACTCGCGTACGTG 2119
Qy 704 LysAlaProGluGly 708
Db 2120 GTCCCGCAGGCGCGC 2134
RESULT 10
US-09-514-340-3
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE D
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs

```

;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3

Alignment Scores:
Pred. No.: 1,24e-19 Length: 2539
Score: 297.00 Matches: 155
Percent Similarity: 31.03% Conservative: 70
Best Local Similarity: 21.38% Mismatches: 224
Query Match: 7.39% Indels: 276
DB: 4 Gaps: 22

US-09-806-767-2 (1-775) x US-09-514-340-3 (1-2539)
Qy 36 SerGluThrAlaLysThrPheAlaSer---LysPheAspTrpHisLeuSerPheLeuGln 54
Db 632 GCGGACGCGCGCGGTGTCGCGAGCGGCAAGCTCGACCGGGGCTCTTCGACATCACC 691
Qy 55 GluAlaValLeuGlyValGluGluGluGluGluGluGluGluGluGluGluGluGlu 74
Db 692 GAACCTCGCAAGCGCGGACCGCGCAACTCCAGAAACAGGAGTGAAGTCACTCGTGGC 751
Qy 75 TyrGlySerAlaIleGluGlyPheAlaAlaGln----- 85
Db 752 TACCAGGCGCGCGACGCGCGCGCGCGAGCTCCGCGAGCGCGCAACTCCGCGCGG 811
Qy 86 ---LeuThrGluSerGluAlaGluLeuArgTyrSerProGluValValAlaValArg 104
Db 812 ACCCTGACCTCCCTCAACGCGGACGCGGTGGCGACCCCGCACGAGCGGTCCGAGCTG 871
Qy 105 ProAspHisValLeuGlnValGlnThrThrTyrSer---TyrLysPheLeuGlyLeuAsp 123
Db 872 TGGGACGCGGTCAACACGCGGACCGGACCGCTCCGCGCATCGCCACGCTCGGTGGAC 931
Qy 124 Gly-----PheGlyAsnSerGlyValTrp 131
Db 932 GGGGTCCGACGCGCGCCCTCGACACGTCCTCGGCGAGATCGCGCCGCCCAAGCGGTG 991
Qy 132 SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTrpPro 151
Db 992 TCCGCGCGCTACGCGGCAAGGCGGTGAGATCCCGTCTCGACACCGGTGTC----- 1045
Qy 152 GluSerProSerPheAspThrGlyMetProSerIleProArgLysTrpLysGlyIle 171
Db 1046 -----GACACAGCATCCGCGACCTG-----AAGGCGCG 1075
Qy 172 CysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGlyAlaArg 191
Db 1076 GTGACCGGTCCAAGAACTTCACGCGCGCGCC----- 1108
Qy 192 PhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArg 211
Db 1108 ----- 1108
Qy 212 GluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGly 231
Db 1109 -----GGCGCGCGGCAAGGTGGCGCACCGCACCCACCTCGCTCGATCGCGCG 1159
Qy 232 GlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMet 251
Db 1160 GGCACGGGC-----GCCAGTCCCAAGGCAAGTACAAAGGGCGTC 1198
Qy 252 AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys---TyrSer 270

```

```

Db 1199 GCACCCGCGCGGATCTCTCAAGCGGAGGTCTCTGACGACTCGGTTCGCGGACGAC 1258
Qy 271 SerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerIle 290
Db 1259 TCCGCGCATCTCTCGCGCATGAGTGGCGCGCGCGAGCGCGCGCATGCTGTCTACCATG 1318
Qy 291 SerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArg 310
Db 1319 AGCTTGGCGCGCATGGACACACACCGAGACCGCTGGAGCGCGCGGTGCGACAAGCTG 1378
Qy 311 AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSer 330
Db 1379 TCCGCGGAGAGCGGTCTCTTCCCATCGCGCGCGCAACGAGCGCGCGCGGTGCGATC 1438
Qy 331 SerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArg 350
Db 1439 GGTTCGCGCGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1480
Qy 351 PheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrPro 370
Db 1480 ----- 1480
Qy 371 GlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAsp 390
Db 1481 -----GTGACGCAAGGAC 1495
Qy 391 LysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMet 410
Db 1496 AAG----- 1498
Qy 411 ValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAla 430
Db 1498 ----- 1498
Qy 431 GlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSerIleAsp 450
Db 1498 ----- 1498
Qy 451 ValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyr 470
Db 1498 ----- 1498
Qy 471 ValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArg 490
Db 1498 ----- 1498
Qy 491 SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer 510
Db 1499 -----CTCGCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1543
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGln--- 529
Db 1544 ATC---AAGCGGAGGTCACTCCCGCTCGGATCGACATCAGCGCGCTCTCTCTCTCTCT 1600
Qy 530 -----AsnLeuGlyProThrGlyLeuProTyrAspSerArg 542
Db 1601 AACGACATCGCGCGAGGTCTGAGGACCGCGCGCGC----- 1639
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1640 -----TACATGACCATCTCCGCGCATGTCGATGGCGACCCCGCACGTCGCGCGCG 1693
Qy 563 AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet 582
Db 1694 GCCCTCTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1750
Qy 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1751 -----ACCGGTCTCCACCAAGGCG 1768
Qy 603 Gly-----ValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619

```

Db 1769 GGCAGGTACACCCCGTTCCAGCAGGGTTCGGCCGGATCCAGGCCGACAAAGCGCTCCAG 1828
Qy 620 ProGlyLeuVal-----TyrAsnIleGln----- 627
Db 1829 CAGACCGTATCCCGCAGACCGCGTCTCGGTAGCTTCGGGTCCAGCAGTGGCGGCACACC 1888
Qy 628 -----ProVal---AspTyrIleThrThrLeuCysThrLeuGlyPheThrArgSer 643
Db 1889 GACGACGCGCGTCCACCAAGCAGCTCACCACCGC---AACCTCGC---ACCCAGGAC 1942
Qy 644 AspIleLeuAlaIleThrHisLeuValSerCysAsnGlyIleLeuArgLysAsnPro 663
Db 1943 GTCAGCTGAAGCTGAGCTCAGCGCCACCGACCCCAAGCGCAAGCGCGCCCGCGCGC 2002
Qy 664 GlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGlu 683
Db 2003 TTCTTCACGTGGCGCCACACCGTACCGTCCCGCGCGCGCGCGCGCTCCGTCGAC 2062
Qy 684 MetIleThrArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnVal 703
Db 2063 ATG---ACCGCGNACCCCGCTCGCGCGCACGGTGGACGGCGCTACTCGCGGTACGTG 2119
Qy 704 LysAlaProGluGly 708
Db 2120 GTCGCCACGGCGCGC 2134

RESULT 11

US-09-000-016-1
; Sequence 1, Application US/09000016
; Patent No. 6143541

GENERAL INFORMATION:

APPLICANT: Akira ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000.016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914

ORIGINAL SOURCE:

ORGANISM: Streptomyces antibioticus
STRAIN:

FEATURE:

NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1

Alignment Scores:

Pred. No.: 1.48e-19 Length: 2809
Score: 297.00 Matches: 155
Percent Similarity: 31.03% Conservative: 70
Best Local Similarity: 21.38% Mismatches: 224
Query Match: 7.39% Indels: 276
DB: 3 Gaps: 22

US-09-806-767-2 (1-775) x US-09-000-016-1 (1-2809)

Qy 36 SerGluThrAlaLysThrPheAlaSer---LysPheAspTrpHisLeuSerPheLeuGln 54
Db 632 GCGGACGCGCGCGCGCTGGTCCGAGCGGCAAGCTCGACCGCGGCTCTTCGACATCACC 691
Qy 55 GluAlaValLeuGlyValGluGluGluGluGluGluGluGluGluGluGluGluGlu 74
Db 692 GAACTCGGCAAGCGCGGACCCCGCACTCCCGAAGAACAGGACTGAAGGTCATCTCGCGC 751
Qy 75 TyrGlySerAlaIleGluGlyPheAlaAlaGln----- 85
Db 752 TACAGGCGCGCGACG 811
Qy 86 ---LeuThrGluSerGluAlaGluLeuArgTyrSerProGluValAlaValArg 104
Db 812 ACCCTGACGTCCTTGACGCGGACGCGGTGGCGACCCCGCACGAGGACGCGTCCGAGCTG 871
Qy 105 ProAspHisValLeuGluValGlnThrThrTyrSer---TyrLysPheLeuGlyLeuAsp 123
Db 872 TGGGACGCGCTCACCACGCGGACCGCGCCCTCCGCGCATCGCCCGACGCTCGGTGTCAC 931
Qy 124 Gly-----PheGlyAsnSerGlyValTyr 131
Db 932 GGGGTCCGACGCGCGCCCTCGACACGTCCTCGGCGAGATCGCGCGCGCGCGCGCGTGG 991
Qy 132 SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTyrPro 151
Db 992 TCCGCGCGCTACGACGCGGCGGTGAAGATCGCGTCTTGACACCGGTGTC----- 1045
Qy 152 GluSerProSerPheAspThrGlyMetProSerIleProArgLysTyrLysGlyIle 171
Db 1046 -----GACACGAGCATCCGACCTG-----AAGGCGCGG 1075
Qy 172 CysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGlyAlaArg 191
Db 1076 GTGACCGCGTCCAGAGACTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1108
Qy 192 PhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArg 211
Db 1108 ----- 1108
Qy 212 GluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGly 231
Db 1109 -----GGCGCGCGGACAGGTGGCGCGCACCGCGCGCGCGCGCGCGCGCGCG 1159
Qy 232 GlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMet 251
Db 1160 GGCACGCGGCG-----GCCAGTCCAAAGGCGAAGTACAGGCGCGCTC 1198
Qy 252 AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys---TyrSer 270

Db 1199 GCACCCGGCGCGGATCTCAACGGCAAGTCTCTGACAGCTCCGGTTTCGGCGCAGC 1258
Qy 271 SerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeu 290
Db 1259 TCCGGATCTCCCGCATGGAGTGGCGCGCGCGCAGGGCCCGCATGTCGTCACATG 1318
Qy 291 SerLeuGlyGlyPheProIleProLeuTyAspAspThrIleAlaIleGlyThrPheArg 310
Db 1319 AGCTGGCGCGCATGGACACACCGGACCGACCGCTGGAGCGCGGTCGACAAAGCTG 1378
Qy 311 AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnGlyProIleGluSer 330
Db 1379 TCCGCCAGAGGGCGTCTGTCGTCATCGCGCGCGCAACAGGCGCGGAGTCGATC 1438
Qy 331 SerValAlaAsnThrAlaProTyrValSerThrIleGlyAlaGlyIleLeuAspArg 350
Db 1439 GGTTCGCGCGCGCAGCGCGCGCGCTCACGTCGGCGCC----- 1480
Qy 351 PheProAlaValArgLeuAlaAsnGlyLysLeuLeuTyrglyGluSerLeuTyPro 370
Db 1480 ----- 1480
Qy 371 GlyLysGlyLysAsnAlaGlyArgGluValGluValIleTyValThrGlyGlyAsp 390
Db 1481 -----GTCCAGCAGCAAGGAC 1495
Qy 391 LysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMet 410
Db 1496 AAG----- 1498
Qy 411 ValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAla 430
Db 1498 ----- 1498
Qy 431 GlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSerIleAsp 450
Db 1498 ----- 1498
Qy 451 ValHisLeuLeuProAlaThrLeuIleGlyTyThrGluSerValLeuLeuLysAlaTy 470
Db 1498 ----- 1498
Qy 471 ValAsnAlaThrValLysProLysAlaArgIlePheGlyThrValIleGlyArg 490
Db 1498 ----- 1498
Qy 491 SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer 510
Db 1499 -----CTCGCGGACTTCTCTCCACCGCGCGCGCTCGCGCAGCGCGCC 1543
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleAlaAlaIleProGln--- 529
Db 1544 ATC---AAGCGGAGCGTCACGCTCCCGCGGTGACATCACGCGCGCTCGCGGAGGCG 1600
Qy 530 -----AsnLeuGlyProThrGlyLeuProTyrAspSerArgArg 542
Db 1601 AACGACATCGCGCAGGAGTCTGAGGACCGCGCGCG----- 1639
Qy 543 ValAsnPheThrValMetSerGlyTyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1640 -----TACATGACCATCTCGGCACGTCGATGGCAGCCCGCATCGCGCGCGCGCG 1693
Qy 563 AlaLeuIleArgSerAlaTyProAsnTyrSerProAlaAlaIleLysSerAlaLeuMet 582
Db 1694 GCCTCTCTGAAGCAGCAGCAGCCCGACTGGACCTCGCGCAGTGAAGGCGCGCTC--- 1750
Qy 583 ThrThrAlaAspLeuTyAspArgGlnGlyAlaIleLysAspGlyAsnLysProAla 602
Db 1751 -----ACCGCTCCACCAAGGCG 1768
Qy 603 Gly-----ValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
Db 1769 GGCAAGTACACCGCTTCGAGCAGGGTTCGGCGCGGATCCAGCGCGCAGCAAGCGCTCCAG 1828

Qy 620 ProGlyLeuVal-----TyrAsnIleGln----- 627
Db 1829 CAGACCGTATCGCGCAGCCCGCTCGGTCTCGGTAGCTTCGGCTGCCAGTGGCGCAGCACC 1888
Qy 628 -----ProVal---AspTyrIleThrTyLeuCysThrLeuGlyPheThrArgSer 643
Db 1889 GACGACGAGCGCGTCAACAGCAGCTACCTACCGC---AACCTCGGC---ACCCAGGAC 1942
Qy 644 AspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnPro 663
Db 1943 GTCAGCTGAGCTGACGTGCGACCGCCACCGACCCCAAGGCAAGGCGCGCGCGCGG 2002
Qy 664 GlyPheSerLeuAsnTyProSerIleAlaValIlePheLysArgGlyLysThrGlu 683
Db 2003 TTCCTTCAGCTGGCGCCACCGACCGTACCGTCCCGCGCGCGCGCAGCGCTCGCTCGAC 2062
Qy 684 MetIleThrArgArgValThrAsnValGlySerProAsnSerIleTySerValAsnVal 703
Db 2063 ATG---ACCGCGCAGACCGCGCTCGCGCGCAGCGTGCAGCGCGCTACTCGGCGTACGTG 2119
Qy 704 LysAlaProGlyGly 708
Db 2120 GTCGCCACCGCGCGC 2134
RESULT 12
US-09-514-340-1
Sequence 1, Application US/09514340
Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE D
ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <Unknown>
FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1

Alignment Scores:
Pred. No.: 1,48e-19 Length: 2809
Score: 297.00 Matches: 155
Percent Similarity: 31.03% Conservative: 70
Best Local Similarity: 21.38% Mismatches: 224
Query Match: 7.39% Indels: 276
DB: 4 Gaps: 22

US-09-806-767-2 (1-775) x US-09-514-340-1 (1-2809)
QY 36 SerGluThrAlaLysThrPheAlaSer---LysPheAspTrpHisLeuSerPheLeuGln 54
DB 632 GCGAGCGCCCGCGGCTGTGTCGCGAGCGGCAAGCTCGACCGGGCTCTTCGACATCACC 691
QY 55 GluAlaValLeuGlyValGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 74
DB 692 GAATCGGCAAGCGCGGACCCCAACTCCAGAAACAGGACTGAGGTCTATCTGTCGGC 751
QY 75 TyrGlySerAlaIleGluGlyPheAlaAlaGln----- 85
DB 752 TACCAGGGCGCGCAGCGGCGCCAGAGCGGAGGTCGCGAGCGGGCGAACTCCGCGCGG 811
QY 86 ---LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArg 104
DB 812 ACCCTGACGTCCCTGAACGGCGGAGCGGTGCGGACCCCGCACGCGCGTCCGAGCTG 871
QY 105 ProAspHisValLeuGlnValGlnThrTyrSer---TyrLysPheLeuGlyLeuAsp 123
DB 872 TGGGACCCCGTACCAGCGGCGAGCGGACCGGCTCGGATCCGCCACGCTCTGGCTGGAC 931
QY 124 Gly-----PheGlyAsnSerGlyValTrp 131
DB 932 GGGGTCCGAGGGCGCCCTCGACACGTCGTCGGGCGAGATCGGCGCCCAAGGGGTGG 991
QY 132 SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTrpPro 151
DB 992 TCCCGCGGTACACGCGCAAGGGCGTGAAGATCGCGTCTCGACACCGGTGTC----- 1045
QY 152 GluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyIle 171
DB 1046 -----GACACGACCATCCGGAACCTG-----AAGGGCCGG 1075
QY 172 CysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLysLeuIleGlyAlaArg 191
DB 1076 GTGACCGCGTCCAGAACTTACCGCGCGCGCC----- 1108
QY 192 PhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArg 211
DB 1108 ----- 1108
QY 212 GluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGly 231
DB 1109 -----GGCGCGCGCGACAGGTGGGCCACCGCACCCACGCTCGCTCGATCGCGGG 1159
QY 232 GlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMet 251
DB 1160 GGCACGCGG-----GCCAGTCCAGGGCGAAGTACAAGGGCGTC 1198
QY 252 AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys---TyrSer 270
DB 1199 GCACCGCGCGCGGATCCTCAACGGCAAGGTCTCTCGACGACTCCGCTTCGCGCGACG 1258
QY 271 SerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeu 290

```

```

DB 1259 TCCGCGCATCTCTCGCGCATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1318
QY 291 SerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArg 310
DB 1319 AGCTGGCGCGCATGGACACACCGGAGACCGACCGCTCGAGGCGGGTTCGACAAGCTG 1378
QY 311 AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSer 330
DB 1379 TCCGCCGAGAAGGGCGTCTTCGCCATCGCGCGCGCAACGAGGCGCGCGAGTGCATC 1438
QY 331 SerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArg 350
DB 1439 GGTTCGCGCGCGCAGCGCGCGCGCGCGCTCACGCTCGCGCGC----- 1480
QY 351 PheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrPro 370
DB 1480 ----- 1480
QY 371 GlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAsp 390
DB 1481 -----GTGACGACCAAGGAC 1495
QY 391 LysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMet 410
DB 1496 AAG----- 1498
QY 411 ValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAla 430
DB 1498 ----- 1498
QY 431 GlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSerIleAsp 450
DB 1498 ----- 1498
QY 451 ValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyr 470
DB 1498 ----- 1498
QY 471 ValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArg 490
DB 1498 ----- 1498
QY 491 SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer 510
DB 1499 -----CTCGCGCGACTTCTCTCTCCACCGCGCCCGCCCTCGGCGCGCGCGCC 1543
QY 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGln--- 529
DB 1544 ATC---AAGCCGAGCGTCAACGCTCCCGCGGTGGACATCACGCGCGCTCGGCGAGGGC 1600
QY 530 -----AsnLeuGlyProThrGlyLeuProTyrAspSerArgArg 542
DB 1601 AACGACATCGCGCAGGAGTGGTGGAGGACCGCGCGC----- 1639
QY 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
DB 1640 -----TACATGACCATCTCCGCGACGTCGATGCGACCGCGCGCGCGCGCGCGCG 1693
QY 563 AlaIleIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet 582
DB 1694 GCCCTCTTGAAGACGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1750
QY 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
DB 1751 -----ACCGGCTCCACCAAGGGC 1768
QY 603 Gly-----ValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
DB 1769 GGCAGAGTACACCCCGTTCGAGCAGGTTTCGGCGCGGATCCAGGCGCGACAGGCGCTCC 1828
QY 620 ProGlyLeuVal-----TyrAsnIleGln----- 627

```

Db 1829 CAGACGGTATCGCCGACCCCGGTCTCGGTGAGCTTCGGGTCCAGCAGTGGCCGACACC 1888
 Qy 628 -----ProVal---AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSer 643
 Db 1889 GACGACGAGCGGTCAACAGCAGCTGACCTACCGC---AACCTCGGC---ACCAGGAC 1942
 Qy 644 AspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnPro 663
 Db 1943 GTCACGCTGAAGCTGACGTCGACCGCCACCGACCCGACCCCAAGGCGGCGCCGCGGC 2002
 Qy 664 GlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrGlu 683
 Db 2003 TTCCTTCAGCTGGGCGCCACACGGTGACCGTCCCGGCGGCGGCGGCGCTCCGTGCAC 2062
 Qy 684 MetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnVal 703
 Db 2063 ATG---ACGCGCAGACCGCGTCCGCGCGCACGGTGACGCGCGGTACTCGCGGTACGTG 2119
 Qy 704 LysAlaProGluGly 708
 Db 2120 GTCGCGCAGCGCGGC 2134

RESULT 13

US-09-255-502-1
 ; Sequence 1, Application US/09255502
 ; Patent No. 6218165
 ; GENERAL INFORMATION:
 ; APPLICANT: Estell, David
 ; APPLICANT: Harding, Fiona
 ; TITLE OF INVENTION: Mutant Proteins Having Lower Allergenic Responses in
 ; TITLE OF INVENTION: Humans and Methods for Constructing, Identifying and
 ; TITLE OF INVENTION: Producing Such Proteins
 ; FILE REFERENCE: GC 527-D2
 ; CURRENT APPLICATION NUMBER: US/09/255,502
 ; CURRENT FILING DATE: 1998-02-23
 ; PRIOR APPLICATION NUMBER: 09/060,872
 ; PRIOR FILING DATE: 1998-04-15
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1494
 ; TYPE: DNA
 ; ORGANISM: Bacillus amyloliquefaciens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (96)..(1244)
 ; NAME/KEY: mat_peptide
 ; LOCATION: (417)..(1241)
 US-09-255-502-1

Alignment Scores:

Pred. No.: 1.89e-17 Length: 1494
 Score: 271.50 Matches: 126
 Percent Similarity: 30.73% Conservative: 63
 Best Local Similarity: 20.49% Mismatches: 177
 Query Match: 6.76% Indels: 249
 DB: 3 Gaps: 17

US-09-806-767-2 (1-775) x US-09-255-502-1 (1-1494)

Qy 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLys 25
 Db 129 TTGCTTTAGCGTTAATCTTTACGATGGCGGTTCGCGACACATCTCT----- 176
 Qy 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLys 45
 Db 177 -----GCCGCGCGCAGGGAATCAACGGGGGAAGAATAATATTCTCGG 224
 Qy 46 PheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 65
 Db 225 TTTAAACAGACAAATGACGACGATGAGCGCGCTAAGAGAGAGATGTCATTCTGAAAAA 284
 Qy 66 GluProSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGln 85

Db 285 GCGCGGAAGTGCACAAAGCAATTCAAATAT-----GTAGCGCAGCTTCAGCTACA 335
 Qy 86 LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArgPro 105
 Db 336 TTTAAACGAAAAGCTCTAAAGAATTGAAAAAGACCCGAGCGTCCGTTACGTTGAAGAA 395
 Qy 106 AspHisValLeuGlnValGlnThrThrTyrSerTyrLysPhe---LeuGlyLeuAspGly 124
 Db 396 GATCAGCTAGCACAT-----GCGTACGCGCAGTCCGTGCCTTACGCGTATCACAA 446
 Qy 125 PheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyVal 144
 Db 447 ATTAAAGCCCTGCTCTGCACCTCAAGGCTACACTGGATCAAAATGTTTAAAGTAGCGTT 506
 Qy 145 LeuAspThrGlyValTrpProGluSerProSerPheAspThrGlyMetProSerIle 164
 Db 507 ATCGACGCGGTATCATCTCTCTCATCTCTGATTTAAAGTAGCAGGCGGAGCCGACATG 566
 Qy 165 ProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsn 184
 Db 566 ----- 566
 Qy 185 ArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGlu 204
 Db 566 ----- 566
 Qy 205 GluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThr 224
 Db 567 -----GTTCTCTTCTGAAACAAATCTTTCCAGACACAACTCTCACGGAAC 614
 Qy 225 HisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGly 244
 Db 615 CAGGTGCGCGCACAGTTCCG-----GCTCTTAATAACTCA 650
 Qy 245 AlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrp 264
 Db 651 ATCGGTGTATTA---GGCGTTCGCCAAGCGCATCACTTTCAGCTGTTAAAGTTCTCGGT 707
 Qy 265 PheAsnGlyCys---TyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAsp 283
 Db 708 GCTGACGGTTCGCGCAATACAGCTGGATCAATACGGAATCGAGTGGCGGCGATCAAC 767
 Qy 284 LysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThr 303
 Db 768 AATATGGAGCTTATTAAACATGAGCTTCGCGGA-----CCTTCGTGTTCTCGTCT 818
 Qy 304 IleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGly 323
 Db 819 TTTAAAGCGCGCAGTTGATAAAGCCGTTGCATCCGCGTCCGATGCTGCGGCGACCGGT 878
 Qy 324 AsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGly 343
 Db 879 AACGAAGGCACCTCCGCGCAGCTCA-----AGCACAGTGGGC 914
 Qy 344 AlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeu 363
 Db 914 ----- 914
 Qy 364 TyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluVal 383
 Db 915 -----TACCCTGGTAAA----- 926
 Qy 384 IleTyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArg 403
 Db 926 ----- 926
 Qy 404 GluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLys 423
 Db 926 ----- 926
 Qy 424 GlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsn 443

Db 926 ----- 926
Qy 444 GlnGluAspSerIleAspValHisLeuLeuProAlaThrLeu---ileGlyTyrThr 462
Db 927 -----TACCCTTCTGTCAATTCAGTAGCGGTGT 956
Qy 463 GluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIle 482
Db 957 GACAGC----- 962
Qy 483 PheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArg 502
Db 963 -----AGCAACCAAGAGCATCTTTCTCAAGCGTA 992
Qy 503 GlyProSerIleuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsn 522
Db 993 GGACCTGAGCTT-----GATGTCATGGCACCCTGGCGTATCT 1028
Qy 523 IleIleAlaAlaTProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArg 542
Db 1029 ATCAAGACGCTTCTCGAAGC----- 1052
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1053 ---AAATACGGGGCGGTACAAAGGTACGTCAATGGCATCTCCGACGTTGCCGAGCGGCT 1109
Qy 563 AlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaAlaIleLysSerAlaLeuMet 582
Db 1110 GCATTGATCTTTCTAAGCACCCGAACTGGACAAACACTCAAGTCGCGACGCTTTAGAA 1169
Qy 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1170 AACACCACT-----ACAAAACCTGGT 1190
Qy 603 GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAla 617
Db 1191 GATTCTTCTACTGGAAGGCTGATCAAGCTACAGCGCA 1235
RESULT 14
5472855-1
; Patent No. 5472855
; APPLICANT: CARTER, PAUL J.; WELLS, JAMES A.
; TITLE OF INVENTION: SUBSTRATE ASSISTED CATALYSIS
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION NUMBER: US/08/287,964
; FILING DATE: 22-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 90,902
; FILING DATE: 12-JUL-1993
; APPLICATION NUMBER: 823,039
; FILING DATE: 14-JAN-1992
; APPLICATION NUMBER: 35,652
; FILING DATE: 06-APR-1987
; APPLICATION NUMBER: 334,081
; FILING DATE: 04-APR-1989
; APPLICATION NUMBER: 127,134
; FILING DATE: 01-DEC-1987
; APPLICATION NUMBER: 846,627
; FILING DATE: 01-APR-1986
; APPLICATION NUMBER: 858,594
; FILING DATE: 30-APR-1986
; APPLICATION NUMBER: 614,612
; FILING DATE: 29-MAY-1984
; APPLICATION NUMBER: 614,615
; FILING DATE: 29-MAY-1984
; APPLICATION NUMBER: 614,617
; FILING DATE: 29-MAY-1984
; APPLICATION NUMBER: 614,491
; FILING DATE: 29-MAY-1984
; SEQ ID NO: 1:
; LENGTH: 1497
5472855-1

Alignment Scores:
Pred. No.: 1.88e-17 Length: 1497
Score: 271.50 Matches: 126
Percent Similarity: 20.73% Conservative: 63
Best Local Similarity: 20.49% Mismatches: 177
Query Match: 6.76% Indels: 249
DB: 6 Gaps: 17
US-09-806-767-2 (1-775) x 5472855-1 (1-1497)
Qy 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLys 25
Db 129 TTTGCTTTAGCGTTAAATCTTTTACGATGCGCTTCGGCAGCACATCCTCT- 176
Qy 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLys 45
Db 177 -----GCCAGCGGCGCAGGGAATCAACGGGGAAGAAATATATTTGTCGGG 224
Qy 46 PheAspTyrHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 65
Db 225 TTTAAACAGACAATGAGCAGCATGCGCGCTAAGAGAGAGATGTCATTTCTGAAAAA 284
Qy 66 GluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGln 85
Db 285 GCGCGGAAAGTGCAAAAGCAATTCAAATAT-----GTAGACGCGAGCTTCAGCTACA 335
Qy 86 LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArgPro 105
Db 336 TTAACGAAAAAGCTGTAAAGAAATGAAAAAGACCCGAGCGCTCGTTAGTTGAAGAA 395
Qy 106 AspHisValLeuGlnValGlnThrTyrSerTyrLysPhe---LeuGlyLeuAspGly 124
Db 336 GATCAGGTAGCACAT-----GCGTACGCGCAGTCCGCTTACGGGGTATCACAA 446
Qy 125 PheGlyAsnSerGlyValTyrSerLysSerArgPheGlyGlnGlyThrIleIleGlyVal 144
Db 447 ATTAAGAGCCCTGCTCTGCACTCTCAAGGCTACACTGATCAATGTTAAAGTAGCGGTT 506
Qy 145 LeuAspThrGlyValTyrProGluSerProSerPheAspThrGlyMetProSerIle 164
Db 507 ATCGACAGCGGTATCGATTCTTCTCATCTGATTAAAGGTAGCAGCGGAGCCAGCATG 566
Qy 165 ProArgLysTyrLysGlyLeuCysGlnGluGlyGluSerPheSerSerSerCysAsn 184
Db 566 ----- 566
Qy 185 ArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGlu 204
Db 566 ----- 566
Qy 205 GluSerProAsnMetProArgGluTyrIleSerAlaArgSerThrGlyHisGlyThr 224
Db 567 -----GTTCCCTTCTGAAACAATCCTTTCCAAGACAACTCTCAGCGAACT 614
Qy 225 HisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGly 244
Db 615 CACGTTGCCGCGCACAGTTGCG-----GCTCTTAATAACTCA 650
Qy 245 AlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrp 264
Db 651 ATCGGTGTATTA---GGCGTTGCGCCAGCGCATCACTTTACGCTGTAAAAGTTCTCGGT 707
Qy 265 PheAsnGlyCys---TyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAsp 283
Db 708 GCTGACGTTCCGGCCCAATACAGCTGGATCATTAACGGAATCGAGTGGCGCATCGCAAC 767
Qy 284 LysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThr 303
Db 768 AATATGACGCTTATTAAACATGAGCTCGCGGA-----CCTTCTGGTCTGCTGCT 818
Qy 304 IleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGly 323
Db ----- 323

```

Db      819  TTTAAAGCGCAGTTGATAAGCCGTTGTCATCCGGCTCGTAGTCGTTGCCGACGCCGT 878
Qy      324  AsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGly 343
Db      879  AACGAAGCAGCTTCCGGCAGCTCA-----AGCACAGTGGGC 914
Qy      344  AlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeu 363
Db      914  ----- 914
Qy      364  TyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluVal 383
Db      915  -----TACCTGTGTA----- 926
Qy      384  IleTyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArg 403
Db      926  ----- 926
Qy      404  GluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLys 423
Db      926  ----- 926
Qy      424  GlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluLeuAsn 443
Db      926  ----- 926
Qy      444  GlnGluGluAspSerIleAspValHisLeuLeuProAlaThrLeu---IleGlyTyrThr 462
Db      927  -----TACCCTTCGTCATGTCAGTAGCGCTGTT 956
Qy      463  GluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIle 482
Db      957  GACAGC----- 962
Qy      483  PheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArg 502
Db      963  -----AGCAACCAAGAGCATCTTTCTCAAGCGTA 992
Qy      503  GlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsn 522
Db      993  GGACCTGAGCTT-----GATGTCATGGCACCCTGGCGTATCT 1028
Qy      523  IleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArg 542
Db      1029  ATCAAGCAGCAGCTTCTCGAATC----- 1052
Qy      543  ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db      1053  ---AAATACGGGGGTACAAACGGTACGTCATGCGCATCTCCGACCGTTCGGGAGCGGT 1109
Qy      563  AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet 582
Db      1110  GCTTTGATTCTTTCTAAGCAGCCGGAACGACAAACACTCAAGTCGCGACGAGTTAGAA 1169
Qy      583  ThrThraAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db      1170  AACACCACT-----ACAAACTGTGT 1190
Qy      603  GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAla 617
Db      1191  GATTCITTCATGATAAGAGGCTGATCAACAGTACAGCGGCA 1235

RESULT 15
US-08-069-863-1
; Sequence 1, Application US/08069863
; Patent No. 5470733
; GENERAL INFORMATION:
; APPLICANT: BRYAN, Philip N
; APPLICANT: ALEXANDER, Patrick
; APPLICANT: STRAUSSBERG, Susan L
; TITLE OF INVENTION: CALCIUM FREE SUBSTITILISIN MUTANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/069,863
; FILING DATE: 01-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028755-016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 450..1599
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 772..1599
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 450
; OTHER INFORMATION: /note="Amino Acid Val at position
; OTHER INFORMATION: 450 is fMet."
; US-08-069-863-1

Alignment Scores:
Pred. No.: 2,76e-17 Length: 1868
Score: 271.50 Matches: 126
Percent Similarity: 30.73% Conservative: 63
Best Local Similarity: 20.49% Mismatches: 177
Query Match: 6.76% Indels: 249
DB: 1 Gaps: 17

US-09-806-767-2 (1-775) x US-08-069-863-1 (1-1868)
Qy      6  PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLys 25
Db      483  TTTGCTTTAGCGTTAATCTTTACGATCGCGTTCGGCAGCACATCTCT----- 530
Qy      26  GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLys 45
Db      531  -----GCCACGGCGGAGGAAATCAACCGGGAAGAAATATATTTGTCGGG 578
Qy      46  PheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 65
Db      579  TTTAAACAGACATGACGACGATGACGCGCTTAAGAGAAAGATGTTCATTTCTGAAAA 638
Qy      66  GluProSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGln 85
Db      639  GGCGGGAAAGTGCAAAAGCAATTCAATAT-----GTAGACGCGAGCTTCAGCTACA 689
Qy      86  LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGlnValValAlaValArgPro 105
Db      690  TTAACGAAAAAGCTGTAAAGAAATTTGAAAAAGACCCGAGCGCTGCTTACGTTGAAGAA 749
Qy      106  AspHisValLeuGlnValGlnThrThrTyrSerTyrLysPhe---LeuGlyLeuAspGly 124

```

```
Db 750 GATCAGTACACAT-----GCGTACGGCGAGTCGCTCGCTTACGGCGTATCACAA 900
Qy 125 PheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleGlyVal 144
Db 801 ATTAAGCCCTCTCTGCACTCTCAAGGCTACACTGGATCAAAAGTTAAAGTACGGTT 860
Qy 145 LeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetProSerIle 164
Db 861 ATCGACAGCGGTATCGATCTCTCATCTCTGATTAAAGGTAGCAGCGGAGCCAGCATG 920
Qy 165 ProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerCysAsn 184
Db 920 ----- 920
Qy 185 ArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGlu 204
Db 920 ----- 920
Qy 205 GluSerProAsnMetProArgGluTyriIleSerAlaArgAspSerThrGlyHisGlyThr 224
Db 921 -----GTTCTCTTGAACAATACTCTTCCAGACACAACTCTCACGGAAT 968
Qy 225 HisThrAlaSerThrValGlyLysSerValSerMetAlaAsnValLeuGlyAsnGly 244
Db 969 CACGTTGCCGGCACAGTTGG-----GCTCTTAATAACTCA 1004
Qy 245 AlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyriLysValCysTrp 264
Db 1005 ATCGGTGTATTA---GGCGTTGGCCCAAGCGCATCACTTACGCTGTATAAAGTTCTCGGT 1061
Qy 265 PheAsnGlyCys---TyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAsp 283
Db 1062 GCTGACGTTCCGGCCNATACAGCTGATCATTAACGGATTCAGTGGCGGATCGCAAC 1121
Qy 284 LysValAspValLeuSerLeuGlyGlyPheProIleProLeuTyriAspAspThr 303
Db 1122 AATATGACGCTTATAACATGACCTCGCGGA-----CCTTCGTGTTCTGCTGCT 1172
Qy 304 IleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGly 323
Db 1173 TTAAGCGCGCAGTGTATAAAGCGTTGTCATCGCGGTGCTAGTCTGCGGCGAGCGCGT 1232
Qy 324 AsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGly 343
Db 1233 AACGAGGCGACTTCCGGCAGCTCA-----AGCACAGTGGGC 1268
Qy 344 AlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeu 363
Db 1268 ----- 1268
Qy 364 TyrGlyGluSerLeuTyriProGlyLysGlyIleLysAsnAlaGlyArgGluValGluVal 383
Db 1269 -----TACCCTGGTAAA----- 1280
Qy 384 IleTyriValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArg 403
Db 1280 ----- 1280
Qy 404 GluGluIleArgGlyLysMetValIleCysAspArgGlyValAlaAsnGlyArgSerGluLys 423
Db 1280 ----- 1280
Qy 424 GlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsn 443
Db 1280 ----- 1280
Qy 444 GlnGluAspSerIleAspValHisLeuLeuProAlaThrLeu---IleGlyTyriThr 462
Db 1281 -----TACCCTTCTGTCATTCAGTAGCGCGCTGT 1310
Qy 463 GluSerValLeuLeuLysAlaTyriValAlaAsnAlaThrValLysProLysAlaArgIleIle 482
```

```
Db 1311 GACAGC----- 1316
Qy 483 PheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArg 502
Db 1317 -----AGCAACCAAGAGCATCTTTCTCAAGCGTA 1346
Qy 503 GlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsn 522
Db 1347 GGACCTGAGCTT-----GATGTCATGGCACCTGCGGTATCT 1382
Qy 523 IleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyriAspSerArgArg 542
Db 1383 ATCCAAAGCAGGCTTCCTGGAAC----- 1406
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1407 -----AAATACGGGGCGGTACAACGGTACGTCAATGGCATCTCCGACGTTGCGGAGCGGCT 1463
Qy 563 AlaLeuIleArgSerAlaTyriProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet 582
Db 1464 GCITTTGATTCTTTCTAAGCACCGGAACTGGACAAACACTCAAGTCCGCGACGAGTTAGAA 1523
Qy 583 ThrThrAlaAspLeuTyriAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1524 AACACCACT-----ACAAAACTTGGT 1544
Qy 603 GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAla 617
Db 1545 GATTCCTTTCTACTATGGAAAAGGCTGATCAAGTACAGCGCGCA 1589
```

Search completed: February 3, 2004, 04:37:39
Job time : 147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 03:56:42 ; Search time 1276 Seconds
(without alignments)
2237.316 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MFEPKPFLLCIIFFLCSSSS.....NSHNLQVRVRSIVTLXTN 775

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US09806767/runat_02022004_090000_18026/app_query.fasta_1.967
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnnpb -MINMATCH=0.1
-LOFPC=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blcsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09806767 -CGN_1_156=brnat_02022004_090000_18026
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:
14: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
17: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:
18: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description

ALIGNMENTS

RESULT 1

US-09-938-842A-1065
; Sequence 1065, Application US/09938842A
; Patent No. US2002160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepis, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCHIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1065
; LENGTH: 2265
; TYPE: DNA

Sequence 1065, Ap
Sequence 1065, Ap
Sequence 1186, Ap
Sequence 1186, Ap
Sequence 1206, Ap
Sequence 485, App
Sequence 271, App
Sequence 1133, App
Sequence 338, App
Sequence 338, App
Sequence 4334, Ap
Sequence 4, Appli
Sequence 64, Appli
Sequence 64, Appli
Sequence 901, App
Sequence 557, App
Sequence 703, App
Sequence 59, Appli
Sequence 620, App
Sequence 5684, Ap
Sequence 3023, Ap
Sequence 3101, Ap
Sequence 3022, Ap
Sequence 5484, Ap
Sequence 6275, Ap
Sequence 346, App
Sequence 228, App
Sequence 15, Appli
Sequence 11, Appli
Sequence 2, Appli
Sequence 5701, Ap
Sequence 1, Appli
Sequence 3306, Ap
Sequence 1, Appli
Sequence 3844, Ap
Sequence 6266, Ap
Sequence 2469, Ap
Sequence 1414, Ap
Sequence 4, Appli
Sequence 7, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 5, Appli
Sequence 6, Appli
Sequence 1, Appli

Db 823 TCGTTATCTCTTGGTGGTGTCTGCTCCGATTATTCGGAACGATTGGAGCG 882

Qy	309	PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAenGlyPProile	328
Dd	883	TTTTTCGGCTATGAGACAGGTGTTTGGTGCTTGTCTCTCGTAATAAGTGGTCTACT	942
Qy	329	GluSerSerValAlaAsnThrAlaProTripValSerThriLeGlyAlaGlyThrLeuAsp	348
Dd	943	AGAGCTTCCTGTCGAATGTTGCTCCTTGGGTTATGACTGTTGGTGCTGTTTAGAT	1002
Qy	349	ArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrglyGluSerLeu	368
Dd	1003	AGAGATTTCCTGGCTTTTGCAGATCTCGTPAACCGGAACAGCACTTACCCTGTTTCCGCTG	1062
Qy	369	TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGly	388
Dd	1063	TATAGCGGTCTAGAATG-----GGACCGAGCCGCTCGAATGGTTTATATAAAGGG	1116
Qy	389	GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleAArgGly	408
Dd	1117	AATAGTAGTCGAGTAATCTTTGTTTACCTGGTTCGCTTATTCGAGTATTGTTCTGGGG	1176
Qy	409	LysMetValIleCysAspArgGlyValAsnGlyArgSerGlnLysGlyGluAlaVallys	428
Dd	1177	AAGATTCTGTTGTGATAGAGGTGTTAAAGCTAGAGTTGAGAAGAGAGCTGTGGTTAGA	1236
Qy	429	GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrCluIleAsnGlnGluAspSer	448
Dd	1237	GATGCTGCTGGTTTAGGGATGATAATGCCAATCTCTCGAGTGGAGAGAGCTGTGTG	1296
Qy	449	IleaspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLys	468
Dd	1297	CGGATAGTCATTTGCTTCCCGCATCGCTGTAGGAAGAGACTGGTGATTACTTAGG	1356
Qy	469	AlatyrValAsnAlaThrValLysProLysAlaAargIlelepheGlyGlyThrValile	488
Dd	1357	GAGTATGTTAAGTCAGATCTCAAACCAACGCTCTTCTGTGTTTAAAGGAACGGTCTT	1416
Qy	489	GlyArgSerArgAlaproGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaasn	508
Dd	1417	GACGTTAAGCGCTCTCTGTGGTTGCTGCTTTTAGCTCGAGAGTCTCTACTGTTACT	1476
Qy	509	ProSerIleLeuLysProAspMetIleAlaproglyValAsnIleIleAlaAtriPro	528
Dd	1477	CCTGAATCTTGAAGCCTGATGTTATTGGTCTCGAGTTAATATTTGGCTGGTGGTCT	1536
Qy	529	GlnAsnLeuglyProthrGlyLeuproTyraSpSerArgValasnPheThrValMet	548
Dd	1537	GACGCTATTGGTCTCTACTGGTCTTGACAAGACTCTAGGAGACTCAGTTCAACATCATG	1596
Qy	549	SerGlyThrSerMetSerCysProHisValserGlyIleThrAlaLeulleArSeraLa	568
Dd	1597	TCAGGTACGTCATGTCTATGCCACACATCAGTGGTGTAGCGGGTCTTTTTGAAAAGCAGCT	1655
Qy	569	TyrProAsnTrpSerProAlaAlailisysSerAlaleuMetThrThrAlaAspleuTyr	588
Dd	1657	CACCTCGAGTGAGTCCGAGTGCTATCAATCACGCTCTCATGACTACAGCTTACGTTCTT	1716
Qy	589	AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla	606
Dd	1717	GACAACCAACCGCTCTCTCCATGATGCTGCACACACACGCTTACTTAACCATATGCT	1776
Qy	607	IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyraSnIle	626
Dd	1777	CACGGCTCGGGCCATGTAGATCCCAAAAGGCTCTCTCACCGAGCTTGTCTACGACATC	1836
Qy	627	GlnProValAspTyrIleThrTyrLeuCysthrIleuglyPheThrArgSerAspileuLeu	646
Dd	1837	TCAACCGAGGAATACATCAGGTTTTTGTGCTCTAGACTACACAGTCGATCACATTGTT	1896
Qy	647	AlaIleThrHisLys---AsnValSerCysAsnGlyIleLeuAryLysAsnProGlyPhe	665
Dd	1897	CGAATGTGAAGCACCGCTAGCGTTTAACTGCTCGAAGAAGTTC---TCAGATCCTGGT---	1955
Qy	666	SerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetile	685

Db 1237 GATGCTGGTGGTTAGGATGATAATGGCGAATPACTGCTCGAGTGGAGGAGCTTG 1296
Qy 449 IleasValHisLeuLeuProAlaThrLeuIleGlyThrGluSerValLeuLeuLys 468
Db 1297 GCGATAGTCAATGCTCCCGGATCGCTGTAGGGAAGAAGACTGGTATTACTTAG 1356
Qy 469 AlaTyrValAsnAlaThrValLysProLysAlaArgIlePheGlyThrValIle 488
Db 1357 GAGTATGTTAAGTCAGATTCTAACCCACCGCTCTCTGTTTAAAGAACGGTCTT 1416
Qy 489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAla 508
Db 1417 GACGTTAAGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1476
Qy 509 ProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleAlaLysPro 528
Db 1477 CCTGAATCTGAGCGCTGATGTTATGGTCCGAGTAAATATTTGGCTGGTGGT 1536
Qy 529 GlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsnPheThrValMet 548
Db 1537 GACGCTATTGGTCTACTGCTCTGACAGGACTCTAGAGGACTCAGTTCACATCATG 1596
Qy 549 SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAla 568
Db 1597 TCAGTACGTCATGTCATGCCACACATCAGTGGTTAGCGGCTCTTTGAAAGCAGCT 1656
Qy 569 TyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyr 588
Db 1657 CACCTCAGTGGAGTCCGAGTGTATCAATCAGCTCTCATGACTACAGCTTACGTTCT 1716
Qy 589 AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606
Db 1717 GACAAACACCGCTCTCTCCATGATGCTGCGACACAGCGCTACTACCCATGCT 1776
Qy 607 IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIle 626
Db 1777 CAGGCTCGGGCCATGATAGTCCCAAGAGCTCTCTCACCAGTCTTGTCACGACATC 1836
Qy 627 GlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeu 646
Db 1837 TCACCCGAGGAATACATCAGGTTTGTGCTCTCTAGACTACACAGCTCATCATGTT 1896
Qy 647 AlaIleThrHisLys---AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe 665
Db 1897 GCGATTCTGAAGCCAGCTAGCTTAATCTGTCGAGGAAGTTC---TCAGATCTGCTGT 1950
Qy 666 SerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIle 685
Db 1951 CAGCTCAACTRCCCAAGTTCTCGGTTTGTGTT---GGGGTAAAGAGTGTGCGGTAC 2007
Qy 686 ThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAla 705
Db 2008 ACTCGGAAGTAAACAATGTTGTGCGACCAAGCTCGGTTTACAAAGTGACGGTTAATGGA 2067
Qy 706 ProGluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGln 725
Db 2068 GCTCCTAGTGTGGAATCTCTGTTAAACCATCGAAACTTCGTTTAAAGCGTGGGAG 2127
Qy 726 ThrLeuSerTyrArgValTyrPheValLeuLysLysAsnArgGlyGlyLysValAla 745
Db 2128 AAGAGAGGTACACAGTCACTGTTGTAGCAAGAAAGAGTGTAGTATGACGAACAGGCT 2187
Qy 746 SerPheAlaGlnGlyGlnLeuThrTyrValAsnSerHisAsnLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GGTTCATCATCTGGAGCAATCCGAGCAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTTGCA 2244

RESULT 3

US-09-938-842A-1186

; Sequence 1186, Application US/0993842A

; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kresps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1186
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1186

Alignment Scores:
Pred. No.: 1,02e-170 Length: 2295
Score: 1589.00 Matches: 342
Percent Similarity: 60.36% Conservative: 127
Best Local Similarity: 44.02% Mismatches: 268
Query Match: 39.55% Indels: 40
DB: 10 Gaps: 16

US-09-806-767-2 (1-775) x US-09-938-842A-1186 (1-2295)

Qy 7 PheLeuCysIleIlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLysGln 26
Db 31 TTCCTCTCTTTCGCTTATCTCTTCGACGCTCTCAGGCC-----GCGAAG 78
Qy 27 ThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPhe 46
Db 79 ACTTTCATTTTCCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 135
Qy 47 AspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 66
Db 136 CATGGTAT-----AGCACCGAGTTCGCCGAGAA--- 165
Qy 67 ProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeu 86
Db 166 -----TCTCGAATCGTCCAGTTTACACACAGCTTCCATGCTTCTCCGCGGTGTT 219
Qy 87 ThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArgProAsp 106
Db 220 ACTCCAGATGAAGCAGATAATCTCCGTAAACCCAGCAGTCTTCTGCTGTTTTCGAAGAC 279
Qy 107 HisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGly 126
Db 280 CGAGCTCGAGAGCTTCACACACAGCTTCTCTCAATTTCTTGGTTTACAA-----AAC 333
Qy 127 AsnSerGlyValTyrPheLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAsp 146
Db 334 CAAAGAGACTATGTCAGATCTGATACGATCAGACGATCAATCAATTCGTTTTCGAC 393
Qy 147 ThrGlyValTyrProGluSerProSerPheAspAspThrGlyMetProSerIleProArg 166
Db 394 ACCGGAATTTGGCCGAGCGGAGGAGTTCTCAGATCTTAACCTCGGTCCCAATCCAAA 453
Qy 167 LysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLys 186
Db 454 AGGTGGAGAGCGGCTTTCGCAATCCGAGCCAGATTCAGTCTCGGAATCTGAACCGTAAA 513
Qy 187 LeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAlaAsnSerProGluGluSer 206
Db 514 ATTATCGGAGCAAGATTCTTCGTAAAGGACAAACAGCCGCT-----GTATCGGAGGA 567

QY 207 ProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThr 226
DB ATCAACAAACCGGTGAGTTCTATCTCTCGTGCAGCGGATGGACACGGTACTCACT 627
QY 227 AlaserThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly 246
DB TCCTCAACCGCGTGGCGCTCAGCTTTTAAAGCGAGTATGTCGGTTACGCCCTCCGT 687
QY 247 ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValIlyrIlyValCysIrp---Phe 265
DB GTAGCCAAAGGTGTGCTCCAAAGCTCGTATCGCGCCCTACAAAGTCTGTTGGAAAGAT 747
QY 266 AsnGlyCysTyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAspIlyVal 285
DB TCGGGTGTCTCGATTCCGATTCTCGCGCTTTCATCGCGCTGTAGACGCGTGC 807
QY 286 AspValLeuSerLeuSerLeuGly-----GlyPheProIleProLeuTyrAspAsp 302
DB GAGCTTATATCGATTCTCAATCGGTGGTGGAGCGGATTAATTCGCCGTATTACCTCGAT 867
QY 303 ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAla 322
DB CCAATCGCTATAGCTCTGACGCGCGCGCTCGAAGGAATCTCTCTCTCTCTCTCTGCC 927
QY 323 GlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProIrpValSerThrIle 342
DB GGAACCAAGGACCTTAACGGTATGTCTAGTTACGAACCTCGCGCGTGGTAAACCGGT 987
QY 343 GlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGlyIlyLeu 362
DB GGTCTAGTACATCGATCGAATTTCCAGCGGATGCTATTCTCGCGCAGCGACATCGT 1047
QY 363 LeuTyrGlyGluSerLeuTyrProGlyIlyLeuIlyAsnAlaGlyValGluValGlu 382
DB CTCAGAGGAGTGTCTTTACGTGGAGTACCTTTAAAC-----GGTCGTATGTTTCCG 1101
QY 383 ValIleTyrVal---ThrGlyGlyAspIlySerGlySerGluPheCysLeuArgGlySerLeu 401
DB GTGGTTATCCCGTAATCCGNAATGTCATGCTCAGCGTCTATGATGAGACACGCTT 1161
QY 402 ProArgGluLeuIleArgGlyIlyMetValIleCysAspArgGlyValAsnGlyArgSer 421
DB GATCCGAAGCAAGTGAGGGGTAAATAGTAATCTGCGATAGAGAAAGCAGTCCACCGTA 1221
QY 422 GluIlyGlyGluAlaValIlyGluAlaGlyValAlaMetIleLeuAlaAsnThrGlu 441
DB GCCAAAGGATGGTTGTGAAGAAAGCAGGTGGTGTGCGAATGATCTCGTAATGGAGCA 1281
QY 442 IleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyr 461
DB TCTAACCGTGAAGGATAGTCGGAGATGCTCATCTTATCCAGCCTGTCGGCTGGATCA 1341
QY 462 ThrGluSerValLeuLeuIlyAlaTyrValAsnAlaThrValIlyProIlyAlaArgIle 481
DB AACGAAGAGATAGAAATCAAGCATATGCTCTTCATCCGATCCGATCCGATTTCAATT 1401
QY 482 IlePheGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAla 501
DB GATTTACAGAGGAATATAGTTGGGATTAACCGCGCTCCGGTATTGCTCTCTCTCCGT 1461
QY 502 ArgGlyProSerLeuAlaAsnProSerIleLeuIlyProAspMetIleAlaProGlyVal 521
DB AGAGGACCAACCGGTTTAAAGCCCGGAGATTTCTTAAACCGGATTTGATGTCCTCCGAGTT 1521
QY 522 AsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArg 541
DB AACATCTCGCGCATGACAGACGCTGTTGGACCTACAGGTTTGGCTCAGATCCAAAG 1581
QY 542 ArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIle 561
DB AAAACGGAATTCACATTTCTCGGTACTTCAATGCGATGCTCTCAGCTTAGTGGTGG 1641

QY 562 ThrAlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaAlaIleIlySerAlaLeu 581
DB CGCGCGCTTCTCAATCCGCTATCCAGATTCGAGCCCTGCCGTGATACGATCCGCAATG 1701
QY 582 MetThrThrAlaAspLeuTyrAspArgGlnGlyIlySerAlaIleIlyAspGlyAsn----- 599
DB ATGACACGATACCTACCTCGTATGATACTTAACCGCTCGTTCGATCGATGATCCACCGGG 1761
QY 600 LysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnIlySerAlaIleAsn 619
DB AAATCGGCTACGCTTATGATTACGGTTCGGGTCAATTAATTTGGGCGCGGTATGAAT 1821
QY 620 ProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGly 639
DB CCGGCTCTGCTACGATATACTAACTAATGATGATTAATACGTTCTTCTTCTCCATCGG 1881
QY 640 PheThrArgSerAspIleLeuAlaIleThrHisIlyAsnValSerCysAsnGlyIleLeu 659
DB TAGGACCAAGACGATCCAGTGATACAAAGAACACCGGTGAGATGTCCGACGACGAGG 1941
QY 660 ArgIlyAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePhe----- 676
DB AAACCGTCTCCGGG---AATTTGAATTTATCTTCGATCAGCGCGGTGTTCTCTACTAAT 1998
QY 677 LysArgGlyIlyThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsn 696
DB AGAAGAGATTTGGTCAGTAAACCTTTATAAGGACGCGCGCAATGTGCGGACGCTCAG 2056
QY 697 SerIleTyrSerValAsnValIlySalapProGluGlyIlyLeuIlyValIleValAsnProIly 716
DB CGCGTTTATCGCGGAGATAGTCCGAGAGAGAGTACGCGTGACAGTGAACACCACT 2118
QY 717 ArgLeuValPheIlyHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuIly 736
DB AGCGTTGTTTACTTCGCGCTTAAGACGCGGAGCTATGCGGTACAGTCAGCTGTTAAT 2178
QY 737 LysIlyAsn-----ArgGlyGlyIlyValAlaSerPheAlaGlnGlyGlnLeuThrTrp 754
DB ACAAGAAATGTTGTTGCGGAGAAACAGGTCTCTGTTT-----GGTCAGTCACGTTGG 2232
QY 755 ValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerValThr 771
DB TTTGATGTGGAGAAACAGCTG-----GTTCCGAGCCCATCGTGGTGACC 2277
RESULT 4
US-09-938-842A-1186
; Sequence 1186, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1186
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1186
Alignment Scores: 1.02e-170 Length: 2295
Pred. No.:

Score: 1589.00 Matches: 342
Percent Similarity: 60.36% Conservative: 127
Best Local Similarity: 44.02% Mismatches: 268
Query Match: 39.55% Indels: 40
DB: 12 Gaps: 16

US-09-806-767-2 (1-775) x US-09-938-842A-1186 (1-2295)

Qy	7	PheLeuCysAlrlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLysGln	26
Db	31	TTCCTCTCTTTCCGGTTTATCTTTCCAGCTCTCTCAGGC-----GCCAAG	78
Qy	27	ThrTyrlleValGlnLeuHisProAsnSerGluThralalysThrPhealaserylSphe	46
Db	79	ACTTTCAATTTCGTATCGATGGTGATCT---ATGCCTTCTATTTCOCGACGCATAC	135
Qy	47	AspTrpHisIeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGlu	66
Db	136	CATTGGTAT-----AGCACCGAGTTCGCCGAGAA---165	
Qy	67	ProSerSerArgLeuLeuTyrrSerTyrglySerAlaileGluGlyPheAlaalaGlnLeu	86
Db	166	-----TCTCGAATCGTCATGTTTACCACACAGCTTCCTCATGTTCTCCGCGGTGTT	219
Qy	87	ThrGluSerGluAlaGluIleLeuArgTyrrSerProGluValValAlaValAlaArgProAsp	106
Db	220	ACTCCAGATGAAGCAGATAATCTCCGTAAACACCCAGCAGTCTGTGTTTTCGAAGAC	279
Qy	107	HisValLeuGlnValGlnThrThrTyrrSerTyryllysPheLeuGlyLeuAspGlyPheGly	126
Db	280	CGAGTCGAGAGCTTACACACACAGTTCTCTCAATTTCTGTTTTACAA-----AAC	333
Qy	127	AsnSerGlyValTrpSerlysserArgPheGlyGlnglyThrIlelleGlyValLeuAsp	146
Db	334	CAAAAAGGACTATGTCAGAACTCATTAACGATCAGACGTAATCATTTGCGTTTTGAC	393
Qy	147	ThrGlyValTrpProGluSerProserPheAspAspThrGlyMetProSerIleProArg	166
Db	394	ACCGGAATTGGCCGAGCGGAGGAGTTCTCAGATCTTAACCTTCGCTCCAATTCACAAA	453
Qy	167	LysTrpLyssglyleCysGlnGluGlyGluSerPheSerSerSerCysAsnArgLys	186
Db	454	AGGTGAGAGCGGCTTTCGAATCCCGAGCCAGATTCACTCTCGGAATCTTAACCGTAA	513
Qy	187	LeulleGlyAlaArgPhePheilleArgGlyHisArgValAlaAsnSerProGluGluSer	206
Db	514	ATTATCGGACAGATTCTTCGTAAGGACACACAAGCCGCT-----GPAATCGGAGA	567
Qy	207	ProAsnMetProArgGluTyrlleSerAlaArgAspSerThrGlyHisGlyThrHisThr	226
Db	568	ATCAACAAAAACCGTTGAGTTTCTATCTCTCGTAGCCGCGATGACACGGTACTCACT	627
Qy	227	AlaserThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly	246
Db	628	TCCTCAACCCGCCCTGGCGGTCAACGTTTAAAGCGAGTATGTCGGGTACGGCTCCGGT	687
Qy	247	ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyryllysValCysTrp---Phe	265
Db	688	GTAGCCAAAGGTGTGCTCCAAAAGCTCGTATCGCGGCTCAAAAGTCTGTTCGAAGAT	747
Qy	266	AsnGlyCysTyrrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysVal	285
Db	748	TCGGTTTGTCTCGATTCCGATATTCTCGCGCGCTTTTGATGCGGTGTGTAGACAGCGTGC	807
Qy	286	AspValLeuSerLeuSerLeuGly-----GlyPheProIleProLeuTyrrAspAsp	302
Db	808	GACGTTTATTCGATCTCAATCGGTGGTGGAGACGGGATTACTTCGCGGTATTACCTCGAT	867
Qy	303	ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyVileSerValIleCysAlaAla	322
Db	868	CCAATCGCTATAGCTCGTAGCGCGCGCGTGCMAAAGGAATCTTCGCTCTCTCTCTGCC	927
Qy	323	GlyAsnAsnGlyProiledGluSerSerValAlaAsnThrAlaProTrpValSerThrIle	342

928	DB	GGAAA	CGAAGCACTAACGGTATGTCAGTTACGAACCTCGCGCGTGGGTAAACCACCGTT	997
343	QY	GlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGlyLysLeu	362	
988	DB	GGTGCTAGTACAATCGATCGAAATTC	1047	
363	QY	LeuTyrGlyGluSerLeuTyrProGlyLysGlyLysAsnAlaGlyArgGluValGlu	382	
1048	DB	CTCAGAGGAGTGTCTTTACGCTGGAGTACCTTTAAAC-----GTCGTATGTTTCCG	1101	
383	QY	ValIleTyrVal--ThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeu	401	
1102	DB	GTGGTTTATCCCGGTAATCTCGGAATGTCTACAGCGTCTCTATGTATGGAGAACACGCTT	1161	
402	QY	ProArgGluGluLeuArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSer	421	
1162	DB	GATCCGAAGCAAGTAGAGGGGTAAATAGTAATCTGGCATAGAGAAAGCAGTCCACCGCTA	1221	
422	QY	GluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGlu	441	
1222	DB	GCCAAAGGATTGGTTGTGAAGAAGCAGGTGGTTCGGAATGATTTCTGGTAATAGAGCA	1281	
442	QY	IleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyr	461	
1282	DB	TCTAACCGGTGAGGATTAGTCGGAGATGCTCATCTTATTCACGCTGTCGCGTTGGATCA	1341	
462	QY	ThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIle	481	
1342	DB	AACGAAGGAGATAGAAATCAAAGCATATGCTCTTCACATCCGAATCCAATGCTTCAATT	1401	
482	QY	IlePheGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAla	501	
1402	DB	GAITTCAGAGGAACATATAGTTGGGATTAACCCGGCTCCGGTTATGTTCTTTCTCCGGT	1461	
502	QY	ArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyVal	521	
1462	DB	AGAGACCAACCGGTTTAAAGCCCGGAGATTCTTAAACCGGATTGATGCTCCCGGAGTT	1521	
522	QY	AsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArg	541	
1522	DB	AACATCTCGCGCATGCAGACAGCGCTGTGGACCTACAGTTTGCCTCAGATCCAAGG	1581	
542	QY	ArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIle	561	
1582	DB	AAAAACGGAAATCAACATCTCTCCGGTACCTCAATGGCATGTCTCCTCAGTTAGTGGTGC	1641	
562	QY	ThrAlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaAlaIleLysSerAlaLeu	581	
1642	DB	CGCGCGCTTCTCAATCCGCTCATCCAGATTGAGCCCTCCGCTGATACGATGGCAATG	1701	
582	QY	MetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsn-----	599	
1702	DB	ATGACACGACTAACCCTCGTCGATACTCAACCGCTCGTTGATCGATGAATCCACCGGG	1761	
600	QY	LysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn	619	
1762	DB	AAATCGGCTACCGCTTATGATTACCGGTCCGGTCAATTAATTTGGCCCGGGTATGAAT	1821	
620	QY	ProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGly	639	
1822	DB	CCGGTCTTGTCTACGATATAACTAATGATGATTACATTACGTTCTTTGCTCATCGGG	1881	
640	QY	PheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeu	659	
1882	DB	TACGGACCAAGACGATCCAAGTGTATACAGAACACCGGTGAGATGTCGACGACCCAGG	1941	
660	QY	ArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePhe-----	676	
1942	DB	AAACCGTCTCCGGG---AATTTGAATATCTCTCGATACGCGGGTGTTCCTACTAAT	1998	
677	QY	LysArgGlyLysThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsn	696	

Db 1999 AGAAGAGGATGGTGAATAAACTGTTATTAAGACGGCGAGAAATGTCGGGAGGCTGAG 2058
Qy 697 SerileTyrSerValAsnValLysAlaProGluGlyLeuLysValIleValAsnProLys 716
Db 2059 GCGGTTATCGGCGAGATAGAGTCGCCGAGAGGAGTGAGCGTGACAGTGAACCACT 2118
Qy 717 ArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTrpPheValLeuLys 736
Db 2119 AGGCTTGTTACTTCGGCGGTTAAGACGCGGAGCTATCGGTTACAGTGACGGTTAAT 2178
Qy 737 LysLysAsn-----ArgGlyGlyLysValAlaLaserPheAlaGlnGlyGlnLeuThrTrp 754
Db 2179 ACAAGGAATGTGTTGGGAGAAACAGGTGCTGTGTT-----GGTCAGTCACGCTG 2232
Qy 755 ValAsnSerHisAsnLeuMetGlnArgValArgValArgSerProLysSerValThr 771
Db 2233 TTGATGTTGGGAAACAGTG-----GTTGGAGCCCCATCGTGTGACC 2277
RESULT 5
US-10-260-238-1206
; Sequence 1206, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1206
; LENGTH: 2043
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (389)..(389)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-1206
Alignment Scores:
Pred. No.: 1,18e-136 Length: 2043
Score: 1290.50 Matches: 297
Percent Similarity: 55.20% Conservative: 112
Best Local Similarity: 40.08% Mismatches: 245
Query Match: 32.12% Indels: 87
DB: 12 Gaps: 17
US-09-806-767-2 (1-775) x US-10-260-238-1206 (1-2043)
Qy 50 LeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSer 69
Db 16 TTGCTTTCCTCTTGTGTTCCATCTTCATATTACACA----- 54
Qy 70 ArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSer 89
Db 55 -----ACTAGTTCACCGGCTACTGAAATTTTCGATCACT----- 90

Qy 90 GluAlaGluIleLeuArgTyrSerProGluValValAlaValArgProAsp----- 106
Db 91 -----AGGCTAGACACATATATAGTGGTGTGCTCCACCAAACTTC 135
Qy 107 -----HisValLeuGlnValGln 112
Db 136 TCAATTGACATGAGCAACATCAAGCTGAGAGTGGTACAGATCATTTCTCCCACTCTTG 195
Qy 113 ThrThrTyrSerTyrLysPheLeuLeuAspGlyPheGlyAsnSerGlyValTrpSer 132
Db 196 ACNAGCCACACACAGATTTCCTAGGCTGGT-----CTCAGGGAAGGATCTCGGAAA 249
Qy 133 LysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTrpProGlu 152
Db 250 AAAACTAGCATGGTGGAGGTGTGATCATAGTGTCTCTGACACAGGATGATTTCACT 309
Qy 153 SerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyLeuCys 172
Db 310 CATACCTCATTTGATGATGATGGATGAGGAGCCACCCACTAAATGCGCGGTCTCTGC 369
Qy 173 GlnGluGlyGluSerPheSerSerCysAsnArgLysLeuIleGlyValaArgPhe 192
Db 370 AAGTCA-----TCTCTGATGAATNCAACAAGAACTCATTTGGGGGTAGCTCA 417
Qy 193 PheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArgGlu 212
Db 418 TTCATTGGGGTCAAAATCAGCA----- 441
Qy 213 TyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGly 232
Db 442 -----CCTCCACCGGATGATGATGGCCATGGGACACATACCGAAGCAGCGAGCTGC 495
Qy 233 SerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAla 252
Db 496 GGTGTTGTAGATGGTCAAGCGTGTGGCAATGGAAATGGCAGCAGCGTGGCATGGCT 555
Qy 253 ProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAsp 272
Db 556 CCTGTGTGCACACCTTGCTATCTACAGTGTGACGATAAAGTTGCGGTGATCTGAC 615
Qy 273 IleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeu 292
Db 616 ATACTTGTGCGCATGGAGCGGCTATTGCTGATGGTGTGACATCATGTCTATCTCTT 675
Qy 293 GlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMet 312
Db 676 GGTGTCCAGCGAAACCATTTTACATGATATAATAGCCACTGTCATCTTTTTCGCCATG 735
Qy 313 GluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerVal 332
Db 736 AGGAAAGGATCTTTGTAAGTCTTGTGCGAGAAATTTCTGGTCCATCTTCCAGCACTCTA 795
Qy 333 AlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPhePro 352
Db 796 AGCAACAGGACCATGGGTTCTGACTGTTGGTCCAGCAGCATAGATAGACAATGGAA 855
Qy 353 AlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLys 372
Db 856 GCCCTTGTCAAGTAGGTGATGGAGACTTATTTCTGGCGAATCTGCTTATCAACCAT 915
Qy 373 GlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGly 392
Db 916 AATCTTGTAT-----CCCTTAGAGTTAGTACCACAG-----ACTTCCGGT 957
Qy 393 SerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIle 412
Db 958 CAAATATTCTCTCTTT-----CTGAAGATGTTGAGAAAGATTTGTGCC 1005
Qy 413 CysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGly 432
Db 1006 TGTGAGCAC---ACAACCTCATCAGACATTTATGACGCTTCGTCAGGATGCTGGTCCA 1062
Qy 433 ValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSerIleAspValHis 452

QY 232 GlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaAaArgGlyMet 251
Db 439 GGTTCACGGTCCCTGCGCTAACTTTGGTATGCCAATGCCACCGCCCAAGGCATG 498
QY 252 AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsn-----Gly 267
Db 499 GCTATAGTGCACACATTCCTCAAGATATGCTGGAGAGATGAGGAATGTCATCG 558
QY 268 CysTyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspVal 287
Db 559 TGCCTACTCTGCATCTTCGCGGAATGAATGAGSCAATTCAGATGAGTAGATGTC 618
QY 288 LeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGly 307
Db 619 ATATCCCTTTCTCTGGTGGTCTAAAGCCCAACTTACACGAACCCACCTCCCTGGC 678
QY 308 ThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyPro 327
Db 679 GCATTCATGCAATCCCTGAGGAGATAGTCTGTCACCTCCGCTGGAATGATGCCCT 738
QY 328 IleGluSerSerValAlaAsnThrAlaProTyrValSerThrIleGlyAlaGlyThrLeu 347
Db 739 GGCACGTACTGTCTAATACCTCGCACCATGGGTGATACGGTTGGCGCATCAAGCAT 798
QY 348 AspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyr---GlyGlu 366
Db 799 GACCGTGGTTCCAGCTCAGTTGTCTTGGACATAATCGCGAACTACATCGGCACC 858
QY 367 SerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrVal 386
Db 859 TCGTGTATTTCCGCCAA-----AACACAGCGGTTTCATTCTACCTCTAGTGTAC--- 909
QY 387 ThrGlyLysAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIle 406
Db 910 ---GGTGTGATGCTGGTTCTGCTGTGTAATACGGGATGCTTCCAGCAACATGGTC 966
QY 407 ArgGlyLysMetValIleCys-----AspArgGlyValAsnGlyArgSerGluLysGly 424
Db 967 ACCGGAAGATTGTCCTCTGCTATGGCACAAATAACTACCAACCCCATAGTGCAGAA 1026
QY 425 GluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln 444
Db 1027 GCTGCTGTCCAAACAGCGGTGTGTGTGGTGCATCATTTCAATTGCTCCAGAAATATGT 1086
QY 445 GluGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSer 464
Db 1087 GATTTCTTACAGACTTCGCTGATATCTGCCAACGTCGACGATTACCTTCAAGGACACC 1146
QY 465 ValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGly 484
Db 1147 GAACGATCCATTGTTACACAGTCGTGGCGGACCCGGTTCCAGAAATCGACTTCTTA 1206
QY 485 GlyThrValIleGlyArgSer---ArgAlaProGluValAlaGlnPheSerAlaAaGly 503
Db 1207 GGCACTGTGATCAACCAATCACCGTCTCTCCAAAGGTCGACGCCITTCACATCGTGA 1266
QY 504 ProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsnIle 523
Db 1267 CCGAACCGCTTTGCTCCAGAGATCCTCAAGCCGACATGATGTCGCCCGGTGTCGACATC 1326
QY 524 IleAlaAlaTrpProGlnAsnLeuGlyProThr---GlyLeuProTyrAspSerArgArg 542
Db 1327 CTCGCTGCATGGACCGGAGAAATGTCCTTACCATGGCCCAATGTCATCGCAATGGCCG 1386
QY 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1387 GTCAGTTTCACATCATCTCGGACCGTTCATGGCTGCATGACATGACGCGCATCGCC 1446
QY 563 AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet 582
Db 1447 GCTATGCTTAAGGTGGCACACCGGAGTTGGAGCCCGCAATCAAGTCGGCCATGATG 1506
QY 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAsp-----GlyAsnLys 600

Db 1507 ACGACGGCCTCAATGTGGCAATGATGGCAATGCCATCAAGGACATGGCCACCGGCCAG 1566
QY 601 ProAlaGlyValPheAlaIleGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 1567 CGAGCCAGCGGCTTCGAGCTTGGTCTGGCCACGTCGACCCCAACCGTCCCTCGACCT 1626
QY 621 GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPhe 640
Db 1627 GGATTTGGTGAACAACACCCCGCGAGATACATATACATTCCTCTGCGAGCTCGGCTAC 1686
QY 641 ThrArgSerAspIleLeuAlaIleThrHis-----LysAsnValSerCysAsnGlyIle 658
Db 1687 AACAGTACTCAGATCGCGCTCTTCAACAATGACGGTTTCGACGACAGACTGTCGCCAGG 1746
QY 659 LeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArg 678
Db 1747 CCACGAGGTCGCGGGC---GACCTCAACTACCGCGGCTTCGCTGGTGTGTTGCTGAGG 1803
QY 679 GlyLysThrThrGluMetIleThrArgArg-----ValThrAsnValGly---SerPro 695
Db 1804 -----TCGGGAGAGCAGTCCACGAGCGCGCGCTGACCAACGTCGGGGCGACACC 1857
QY 696 AsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsnPro 715
Db 1858 AATGTCTGTATCAACGTTTACCATCCGACCGCGCGCACCACTCTCACGTCGACGCCA 1917
QY 716 LysArgLeuValPheLysHisValAspGlnThrIleSerTyrArgValTrpPheValLeu 735
Db 1918 ACGAGGCTCGCGTTCGACGCGCAGCAGACGCTGCTCCGACGACGAGTGGGATCGATCGTCTGG 2022
QY 736 LysLysLysAsnArgGlyGlyLysValAlaSerPheAlaGln---GlyGlnLeuThrTrp 754
Db 1966 ---ACCGTGTGGCAGGCGGCGACGACGCTGCTCCGACGACGAGTGGGATCGATCGTCTGG 2022
QY 755 ValAsnSerHisAsnLeuMetGlnArgValArgSerProfileSerValThr 771
Db 2023 AGCGATGCCAACAT-----ACGCTGAGGAGCGCCGCTGTCGGCCACC 2064

RESULT 7

US-10-259-165-271
; Sequence 271, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 271
; TYPE: DNA
; LENGTH: 2205
; ORGANISM: Oryza sativa
US-10-259-165-271

Alignment Scores:			
Pred. No.:	2,19e-123	Length:	2205
Score:	1175.00	Matches:	293
Percent Similarity:	52.87%	Conservative:	122
Best Local Similarity:	37.32%	Mismatches:	252
Query Match:	29.24%	Indels:	118
DB:	13	Gaps:	27

US-09-806-767-2 (1-775) x US-10-259-165-271 (1-2205)			
Qy	38	ThrAlaLysThrPheAlaSerLysPheAspTrpHisLeu	---SerPheLeuGlnGluAla 56
		...:::...	...:::...
Db	10	ACTCTATCTTCATTCCTCCGCACACAGCTTGACTGGGTCA	TATCATCATCTTCCCTCTACATC 69
Qy	57	ValLeuGlyValGluGluGluGluGluProSer	----- 68
		...:::...	...:::...
Db	70	GTGTACATGGGGAGAAGAAGCATGATCGTCTGTGTTCACGGCATCGCATCATGAC	129
Qy	69	-----SerArgLeuLeuTySerTyr	75
		...:::...	...:::...
Db	130	ACGCTTACTTCTGTTCTTCGGAGTAAGATGGCGCCATGAAGTCTATAGTGTACAGTTAC	189
Qy	76	GlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluIleLeuArg	95
Db	190	AAACATGGGTTCCTCGATTTCGACGATGCTCACCGAATCCCAAGCTGAGGAACCTTGA	249
Qy	96	TyrSerProGluValValAlaValArgProAspHisValLeuGlnValGlnThrThrTyr	115
Db	250	AGATTACCTGAAGTCAATCAGCGTGAAGCTTAACACTTACCATTCAAGCACAAACGACGTGA	309
Qy	116	SerTyrLysPheLeuGlyLeuAspGlyPheGlyAenSerGlyValTyrSerLysSerArg	135
		...:::...	...:::...
Db	310	AGCTGGGATTTCTCGSCCTTAACTACACGAACTCATCGTCTTCTTAAGAAAGCAAG	359
Qy	136	PheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTyrProGluSerProSer	155
		...:::...	...:::...
Db	370	AACGGTGAAGATGTCATCTCGTGTGCATCTCAGGTATATGCCCGGAATCACGAAGC	429
Qy	156	PheAspAspThrGlyMetProSerIleProArgLysTyrLysGlyIleCysGlnGluGly	175
		...:::...	...:::...
Db	430	TTCCGATGACAATGGCTTACAGTCCCGTGCACGATGGAAGGGCAAGTCCGACCGGC	489
Qy	176	GluSerPheSerSerSerSer--CysAenArgLysLeuIleGlyAlaArgPhePheIle	194
		...:::...	...:::...
Db	490	CGCGGTTCAAGCCACACAGGGTTGCACAGGAAGATCATCGCGTACAGATGGTACTCG	549
Qy	195	ArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProA--gGluTyrIle	214
		...:::...	...:::...
Db	550	GGCGGC-----ATTCGACAGAG-----AACCTGAAGGGCGAGTACATG	588
Qy	215	SerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSer	234
		...:::...	...:::...
Db	589	TCGGCAGGAGACCTCGCGGCACAGCGCAGCAGCTGCGCTCGACGATCTGTCGGCGGCGAG	648
Qy	235	ValSerMetAlaAsnValLeuGlyAenGly-----AlaGlyValAlaArgGlyMet	251
		...:::...	...:::...
Db	649	GTACGGAACTGAGCCACACAGAGGGCGGCCCTGGCTGCGGCGACGGCGCGGTGGG	708
Qy	252	AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAenGly-----	267
		...:::...	...:::...
Db	709	GCGCCGCGCGCGGTGCGGTCTACAAGGTGTCTGG-----GGCCTGCGCGCCAG	762
Qy	268	CysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspVal	287
		...:::...	...:::...
Db	763	TGTGGCGGAGCCGATCTCTCCGCCCATTCGACAGCGCCNTAAACAGCGCGTGGATGTG	822
Qy	288	LeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGly	307
		...:::...	...:::...
Db	823	CTGTCCCTGTGATGAGGCGCGGGC---GAGCATTACGAG-----	861
Qy	308	ThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAenAsnGlyPro	327
		...:::...	...:::...

Qy 669 TyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgArg 688
 Db 1882 CTCCTGGTCAATGTCAGTGCCTAAAGACTCTGTACG-----GTTTGGCGACT 1935
 Qy 589 ValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGly 708
 Db 1936 GTCACCAATGTTGAGAGCAGAACCACTTACCATGCACGGTTGAGGCACCGGTGGT 1995
 Qy 709 IleLysValIleValAsnProLysArgLeuValPhe---LysHisValAspGlnThrLeu 727
 Db 1996 ATGACTATGTCAGTGGACCTTCGGTGCATCACATTCATAGGGGTGTAGTAGAAGTGA 2055
 Qy 728 SerTyrArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAlaSerPhe 747
 Db 2056 ACATTTAAGGTGACATTCACAAACACAGACAGTACAAAGTGGC-----TAC 2103
 Qy 748 AlaGlnGlyGlnLeuThrTrpVal-----AsnSerHisAsnLeuMetGlnArgValArg 765
 Db 2104 ACGTTTGGGAGTTTGACATGCTAGATGTTAACAACACACTCT-----GTGAGA 2151
 Qy 766 SerProIleSerVal 770
 Db 2152 ATTCGATTGCAGTT 2166

RESULT 8

US-10-260-238-1133
 ; Sequence 1133, Application US/10260238
 ; Publication No. US20040016025A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Budworth, Paul R.
 ; APPLICANT: Moughamer, Todd G.
 ; APPLICANT: Briggs, Steven P.
 ; APPLICANT: Cooper, Bret
 ; APPLICANT: Glazebrook, Jane
 ; APPLICANT: Goff, Stephen A.
 ; APPLICANT: Katagiri, Fumiaki
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Provart, Nicholas
 ; APPLICANT: Ricke, Darrell
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
 ; FILE REFERENCE: 60111-NP
 ; CURRENT APPLICATION NUMBER: US/10/260,238
 ; CURRENT FILING DATE: 2002-09-26
 ; PRIOR APPLICATION NUMBER: US 60/325,448
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: US 60/325,277
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: US 60/370,620
 ; PRIOR FILING DATE: 2002-04-04
 ; NUMBER OF SEQ ID NOS: 6077
 ; SEQ ID NO 1133
 ; LENGTH: 2334
 ; TYPE: DNA
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; NAME/KEY: N region
 ; LOCATION: (498)..(498)
 ; OTHER INFORMATION: n = any nucleotide
 ; FEATURE:
 ; NAME/KEY: N region
 ; LOCATION: (548)..(548)
 ; OTHER INFORMATION: n = any nucleotide
 ; FEATURE:
 ; NAME/KEY: N region
 ; LOCATION: (551)..(551)
 ; OTHER INFORMATION: n = any nucleotide
 ; US-10-260-238-1133
 Alignment Scores: 2.12e-121 Length: 2334
 Pred. No.: 1158.00 Matches: 291
 Score:

Percent Similarity: 50.73% Conservative: 127
 Best Local Similarity: 35.32% Mismatches: 256
 Query Match: 28.82% Indels: 150
 DB: 12 Gaps: 26
 US-09-806-767-2 (1-775) x US-10-260-238-1133 (1-2334)
 Qy 16 CysSerSerSerSerGluIleLeuGlnLys----- 25
 Db 61 TGTAGTGAAGCAGGCGAGCGTCTCGTACTCCGAATGGCAGCGCGCGCCGCCACCGCAG 120
 Qy 26 -----GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysPheAla 43
 Db 121 GCCTACCCGAGCTTACATCTGCTCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 174
 Qy 44 SerLysPheAsp-----TrpHisLeuSerPheLeuGlnGluAlaVal 57
 Db 175 -----GACGACGACGGCGACCGCGGTGGACGAGTCTTTCTTCCCGGGA----- 219
 Qy 58 LeuGlyValGluGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGlySer 77
 Db 220 ---GGCCGCGCATGACGACGCGCGCGATCAGCAGGATCATCCGCTCCTACACGAG 276
 Qy 78 AlaIleGluGlyPheAlaGlnLeuThrGluSerGlu---AlaGluIleLeuArgTyr 96
 Db 277 GTGTTGAGAGGCTTCGCGCGAGCTACCGCGCGCGAGCTCGCGCGGTGGTGTCTCCAAG 336
 Qy 97 SerProGluValValAlaValArgProAsp---HisValLeuGlnValGlnThrTyr 115
 Db 337 AAGCCCGGGTTCGTGCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396
 Qy 116 SerTyrLysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrp---SerLys 134
 Db 397 AGCCGCGAGTTCCTCGGCTGACG-----AGAGCGCGCGGTTCGAGGAGCGTCCGC 450
 Qy 135 ArgPheGlyGlnGlyThrIleLeuValLeuAspThrGlyValTrpProGluSerPro 154
 Db 451 GCGTACGCGAAGGGCGTCTGCGGTCTGTGCACACCGCGCTCCANGCGCGCACCT 510
 Qy 155 SerPheAspThrGlyMetProSerIleProArgLysTrpLysGlyLeuGlnGlu 174
 Db 511 TCGTTTCGACGACCGCGCGGTCTCCGCGCGCGCGANGTNGAGGGGTCTCGCGCGTC 570
 Qy 175 GlyGluSerPheSerSerSerCysAsnArgLysLeuIleGlyAlaArgPheIle 194
 Db 571 GCC-----GCCACGCGCGGTGCAACACAGCTCGTCGGGTCAAGTCTCTCGTT 621
 Qy 195 ArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArgLysIle 214
 Db 622 GACGCGCGCGC----- 633
 Qy 215 SerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySer 234
 Db 634 GCGCGCGACGACGACGTCGCGCGCGACACACACCGCGTCCAGCGCGCGCGGAACTC 693
 Qy 235 ValSer---MetAlaAsnValLeuGlyAsnGlyValAlaArgGlyMetAlaPro 253
 Db 694 GTCGCGCGCGCGGTCTCGACCGCGGTGCGCGCGGAAACGCGCGCGCGGATCGCGCC 753
 Qy 254 GlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAsp 273
 Db 754 GCGCGCGACGTCGCGGTGTACAGGTGTCAACCGCTCAGGCTCGGATGACGACGCGGTG 813
 Qy 274 LeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeu 293
 Db 814 CTAGCCGGGTTCGACGAGGCCATGAGGACGAGTTCGACGCTCTCTCGGTCTCGTCGG 873
 Qy 294 GlyPheProIleProLeuTyrAsp---AspThrIleAlaIleGlyThrPheArgAlaMet 312
 Db 874 AGGTGTCGAGTCCCGCTTCGATGAGGACCCCATCGCCATCGCGCGCTTCAGCGCAGT 933
 Qy 313 GluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerVal 332

Qy	221	GlyHisGlyThrHisThrAlaSerThrValGlyGlySerValSerMetAlaAsnVal	240
Db	665	GGCCATGGCTCCCACTCTCTCCACCGCGCGAGATTGTGTCTCGAGTCAGCATC	724
Qy	241	LeuGlyAsnGlyAlaGlyValAlaAargGlyMetAlaProGlyAlaHisIleAlaValTyr	260
Db	725	TTCCGCCAAGGCAACGCGCAGCGCAAGGTGGCTCCCTCGCGCCACAGTGGCGGTAC	784
Qy	261	LysValCysTrp-----PheAsnGlyCysTyrSerSerAspIleLeuAlaAa	276
Db	785	AAAGTCTGCTGGCGCCAGTCAAAAGGCAACGAATGCTACGACGACAGCTCTCGCTGCT	844
Qy	277	IleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPhePro	296
Db	845	TTCCAGCGCTGCCATACACGACGGAGCTGACGTCTATCCGTCCTCTAGGGCGGAGCCT	904
Qy	297	IleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluAArgGlyIle	316
Db	905	ACCTCGTGTTTTAAACGACAGTGCCTATTGGTTCCTTTACGCGCGCTAAGAAGCGGATC	964
Qy	317	SerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAla	336
Db	965	GTGGTGTGTCTCTGCGGAAACTCTGGACCGCAGATAGTACGGTTCCTTAATGTGCGG	1024
Qy	337	ProTrpValSerThrIleGlyValaGlyThrLeuAspAargPheProAlaValAlaArg	356
Db	1025	CCATGGCAGATTACCGTTGGTGGCTAGCACCATGACCGGTGAGTTGTGCAATCTTGTT	1084
Qy	357	LeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsn	376
Db	1085	CTCGGTAATGGAAACA-TTACAGGTTAGATATTAAATCCAGAAATATGGATCATATCT	1143
Qy	377	AlaGlyAArgGluValGluVal-----IleTyrValThrGlyGly	389
Db	1144	CTATCTCAAACTTGAAATGAATCTTGAAATGAGCTTGAATCTTTTTCATTCTGAGG	1203
Qy	390	AspLysGly-----	392
Db	1204	GACAAAGCTTGTCATCCACGCGATTGCCACATGTAAAGTTCTATCCAAATATGGCATCTG	1263
Qy	392	-----	392
Db	1264	TAAATGCAAAAGCTAAGAACGCTTCAGCTTTGGATGGGTGAGTTGGCTTAATCTCTWTAC	1323
Qy	393	-----Ser	393
Db	1324	TTAATTACGCCAATGTTTATGTTAATTTGTAATCTCTCTCTCTTTGTTGTAGA	1383
Qy	394	GluPheCysLeuAArgGlySerLeuProAArgGluGluIleAArgGlyLysMetValIleCys	413
Db	1384	CAATTATGCAAACTTGGATCGCTTGACCTATAAAGACAAAGGAAAGATATTAGTGTGT	1443
Qy	414	AspArgGlyValAsnGlyAArgSerGluLysGlyGluAlaValLysGluAlaGlyVal	433
Db	1444	CTCAGAGGCGACAGCGGAGGCTGGAAAAGGCTCGGCGCGTGGCTTTAGGGGGTGGCAT	1503
Qy	434	AlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeu	453
Db	1504	GGGATGGTCTTGTAGACACTTATTCACCGGAATGATTTACTTCGCTGACCCGCGATGC	1563
Qy	454	LeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAla	473
Db	1564	CTTCCAGCCACACGCTCACTTCTAAGGATAGTTTGTCTGTGTCAAGATATATCAGCAA	1623
Qy	474	ThrVal-----	475
Db	1624	ACCAAGTAGGATTCCTGATTCCACACTTTACCTAAACTTGTTTCAAGTTCGAAAGAAAC	1683
Qy	476	-----LysProCysAlaAargIleIlePheGlyGlyThrValIleGly	489
Db	1684	TTTATTTTTTGATAGGAAGCCATATGCACATTTACTCTCTTCAAGACAGATTTGGGA	1743

QY	490	ArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnPro	509
DB	1744	CTAAACCCAGCGCTGTAAATGGCTTCATTTTCCTAAAAAGGTCCAAGCATAGTGGCTCCT	1803
QY	510	SerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTdpProGln	529
DB	1804	CAGATTCGAAAGCCTGACATAACTGCTCCGGTGTGAGTGTGATCGCTGCCTACACTGGA	1863
QY	530	AsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsnPheThrValMetSer	549
DB	1864	CGAGTCCTCCAAACAAAGCAACAGTTTGATCCTCGACGACTTCGTGTCATGCTATTTC	1923
QY	550	GlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyr	569
DB	1924	GGAACTCTATGCTGTGCTCATATCTCTCGGAATGCCGGGTCTCTCAAAACTCGTTAT	1983
QY	570	ProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAsp	586
DB	1984	CCCTCTGGAGCCCGCAGCTATCCGCTTCGCCCATCATGACTACCGGTAGAATGTATTT	2043
QY	586	-----	586
DB	2044	CGATTCAGAGCTTACACATTTGTTGACAAATTAAGAGTTTAATCTTATTTACCACCTAA	2103
QY	586	-----	586
DB	2104	AACACATTAACATCTACTGCTCCTGGACTTATAACCAAATCTTGTTCTTCCATGGCCAT	2163
QY	586	-----	586
DB	2164	GTTCTGTTTTCTAGTTTTTAGTTGGTGAATATTTGGAAATGAGTTTTCTTGAATCTTTACAGT	2223
QY	587	-----LeuTyrAspArgGlnGlyLysAla	594
DB	2224	TTTTTCACATTGCTGTTCCCTTCAGCAACAATAATGGATGACATTCCTGGACCTATC	2283
QY	595	IleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro	614
DB	2284	CAAAATGCAACCAACATGAAGGCAACACCTTTTCAGTTTGGGCGAGACACGCTCAACCA	2343
QY	615	GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr	634
DB	2344	AATTTAGCTGTGAATCCCGTCTCGTATACGATTTAGGCATCAAGGANTTACCTCAACTTC	2403
QY	635	LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer	654
DB	2404	TTATGCTCACTGGATACATGATCACAATCTCTGTTCTCGGGCAACAACCTTTACT	2463
QY	655	CysAsnGlyIleLeuArgLysAsnProGlyPheSer-----LeuAsnTyrProSer	671
DB	2464	TGTTCA-----AGCCCTTAAATCAGTCTTGTTAACTCAACTATCCTCTCC	2508
QY	672	IleAlaVal---IlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThr	690
DB	2509	ATCACAGTTCCTCAACTTAAACATCAAGCAAAAGTCACC-----GTTTCAAGGACTGTAAA	2562
QY	691	AsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLys	710
DB	2563	AACGTTGCAGCACT---TCGATGTATACCGTCAAGTGAATTAACCCACAGGCTGTATAT	2619
QY	711	ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg	730
DB	2620	GTGCGGTGAAGCCACAGCAGCTTTGAATTTTCAACAAAGTTGGAGAGCAAAAGACGTTAAA	2679
QY	731	ValTrpPheValLeuLysLysLysAsnArgGlyGlyLysValAlaLaserPheAlaGlnGly	750
DB	2680	GTGATCCTCTGTG-----AAAAGTAAGGGAATGTGGCTAAGGCGCTATGTTGGTA	2730
QY	751	GlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal	770
DB	2731	GAATTTGTTGGTCAGACAAGAGCAT-----CGTGTGAAGAGTCCCATTTGGTGTG	2781
QY	771	ThrLeu	772

Db 2782 AAGCTC 2787

RESULT 10

US-09-938-842A-338

Sequence 338, Application US/09938842A

Publication No. US20040009476A9

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

TITLE OF INVENTION: SAME, AND METHODS OF USE

FILE REFERENCE: SRIPI300-3

CURRENT APPLICATION NUMBER: US/09/938,842A

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 338

LENGTH: 2958

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-938-842A-338

Alignment Scores:

Pred. No.: 1.29e-111 Length: 2958

Score: 1074.00 Matches: 259

Percent Similarity: 47.38% Conservative: 102

Best Local Similarity: 33.99% Mismatches: 204

Query Match: 26.73% Indels: 198

DB: 12 Gaps: 14

US-09-806-767-2 (1-775) x US-09-938-842A-338 (1-2958)

Qy 185 ArgLysLeuIleGlyAlaArgPheIleArgGlyHisArgValAla----- 200

Db 572 AGGAACTGATGAGGACGAGTACTCAACAGAGGCTCAGCCGCCGCTAGACATCTC 631

Qy 201 AsnSerProGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThr 220

Db 632 AACTCAAGCTTCGACTCTCCA-----CGTGATCTCGAC 664

Qy 221 GlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnVal 240

Db 665 GGCCATGGCTCCACACTCTCTCCACCGCCGCGGAGATTGTTCTCGGAGTCAGCATC 724

Qy 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyr 260

Db 725 TTCGGCCCAAGCAACGCGCACGGCCAAAGGTGGTCTCCCTCGCGCCAGAGTGGCGGTAC 784

Qy 261 LysValCysTrp-----PheAsnGlyCysTyrSerSerAspIleLeuAlaAla 276

Db 785 AAAGTCTGCTGGCGCCAGCTCAAGGCAACGAATGCTACGACGAGCTCTCGTGCT 844

Qy 277 IleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPhePro 296

Db 845 TTCGACGCTGCATACACGAGCGGAGCTGACGTATATCCGTCTCTCTAGCGCGGAGCCT 904

Qy 297 IleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIle 316

Db 905 ACCTCGTTTAAACGACAGTGTGCTGCTATGTTGTTCTTTCACGCGCTAAGAGCGGATC 964

Qy 317 SerValIleCysAlaAlaGlyAsnGlyProIleGluSerSerValAlaAsnThrAla 336

Db 965 GTGGTGTGTCTCTGCGGAACTCTGGACCGGACAGATAGTACGGTTTCTTAATGTCGG 1024

Qy 337 ProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValValArg 356

Db 1025 CCATGGCAGATTACCGTTGGTGTGTAGCACCATGCGTGGAGTTTGTAGCAATCTTGT 1084

Qy 357 LeuAlaIleGlyLysLeuLeuTyrGlyLeuSerLeuTyrProGlyLysGlyIleLysAsn 376

Db 1085 CTCGGTAATGGAAACA-TTCAAGGTTAGATATTAAGTCCAAGATATGATCATATCT 1143

Qy 377 AlaGlyArgGluValGluVal-----IleTyrValThrGlyGly 389

Db 1144 CTATCTCAATCTTGAATAATCTTGAATGACCTTGAATCTTTTCATCTCGTAGG 1203

Qy 390 AspLysGly----- 392

Db 1204 GACAAGCTTCTCATCCAGCAGTCATGCCACATGCTAAGTCTATCAATATGTCATCTG 1263

Qy 392 ----- 392

Db 1264 TAAATGCAAAAGCTAAGAACGCTTACGCTTTGGATGCGTGAGTTTGGCTAATCTCTTAC 1323

Qy 393 -----Ser 393

Db 1324 TTAATTAGCCCAATGTTTATGTTAATGTTAATGTTAATGTTAATGTTAATGTTAATGTTA 1383

Qy 394 GluPheCysLeuArgGlySerLeuProArgGluLeuArgGlyLysMetValIleCys 413

Db 1384 CAATTATGCAAACTTGGATCGCTTACCCCTATAAAGCAAGGGAAGATATTAGTGTGT 1443

Qy 414 AspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyVal 433

Db 1444 CTCAGAGGCGAAGCGGAGGTCGAAAGGTCGGGCGCTTTAGGGGGTGGGATA 1503

Qy 434 AlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeu 453

Db 1504 GGGATGGTCTTGAGAACACTTATGTCACCGGGAATGATTTACTCGCTGACCCCATGTC 1563

Qy 454 LeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLysAlaTyrValAsnAla 473

Db 1564 CTTCCAGCCACACAGCTCACTTCTAAGATAGTTTCTGTGTCAAGATATATCAGCAA 1623

Qy 474 ThrVal----- 475

Db 1624 ACCAAGTAGGATTCTTGATTCACACTTTACCTAAACTTGTTCAGTTCTGAAAGAAAC 1683

Qy 476 -----LysProLysAlaArgIleIlePheGlyGlyThrValIleGly 489

Db 1684 TTTATTTTGTATGTAGGAGCGGATTGACACATTACTCTTCAAGACAGATTGGGA 1743

Qy 490 ArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnPro 509

Db 1744 CTAAACCCAGCGCTGTAATGGCTTCAATTTCTTCAAAAGTCCAGCATAGTGGCTCT 1803

Qy 510 SerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGln 529

Db 1804 CAGATTTCTGAGCTGACATAACTGCTCTGCTGTGAGTGATGCTGCTGCTACACTGGA 1863

Qy 530 AsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMetSer 549

Db 1864 GCAGTCTCTCCAACAACAGCAAGTTGATCTCTCGACGACTTCTGTTCAATGCTATTTC 1923

Qy 550 GlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyr 569

Db 1924 GGAACCTCTATGCTGTGCTCTCATATCTCGCATTCGGGCTCTCTCAAAACGCTTAT 1983

Qy 570 ProAsnTyrSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAsp----- 586

Db 1984 CTTTCTTGAGCCCGGAGCTATCGCTCTGCCATCATGACTACCGTAAGATGATTT 2043

Qy 586 ----- 586

Db 2044 CGATTCTAGAGCTTACATTTGTGACAAATGGAAGTTTAATCTATTTTACCCTTAA 2103

Qy 586 ----- 586

Db 2104 AACACATTACATCTACTGCTCGGACTTATAACAAAATCTTGTTCCTCCATGGCCAT 2163
Qy 586 -----
Db 2164 GTTCTGTTTCTTAGTTAGTTGGAAATATTGGATGAGTTCTTGTGAATCTTACAGT 2223
Qy 587 -----LeuTyrAspArgGlnGlyLysAla 594
Db 2224 TTTTTCACACTTGTGTTTCCCTTTCAGCAACAATAATGGATGACATTCCTGGACCTATC 2283
Qy 595 IleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db 2284 CAAATGCAACCAACAAAGGCAACACCTTTCAGTTTGGGGGAGGACAGCTCCAAACCA 2343
Qy 615 GlnLysAlaIleAsnProGlyLeuValTyrAsnLysGlnProValAspTyrIleThrTyr 634
Db 2344 AATTAGCTGTGAATCCCGTCTGTATACGATTAGGCATCAAGGATTACCTCAACTTC 2403
Qy 635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer 654
Db 2404 TTATGCTCAGTTGATACATCAATGTCATCAAAATCTCTGTATTCTCGGGCAACAACCTTTACT 2463
Qy 655 CysAsnGlyIleLeuArgLysAsnProGlyPheSer-----LeuAsnTyrProSer 671
Db 2464 TGTTCAC-----AGCCCTAAATCACTCTGTGTAACCTCAACTATCTCTCC 2508
Qy 672 IleAlaVal-----IlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThr 690
Db 2509 ATCAGAGTCCAAACTTAACATCAAGCAAAAGTCACC-----GTTTCAAGGACTGTTTAAA 2562
Qy 691 AsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLys 710
Db 2563 AACGTTGGAGCACCT---TCGATGTATACCTCAAGGTGATATACCCACAGGGTGTAT 2619
Qy 711 ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg 730
Db 2620 GTCGGGTGAAGCAACAGAGCTTGAATTTTCCAAAGTTGAGAGACAAAGAGCTTTAAA 2679
Qy 731 ValTrpPheValLeuLysLysLysAsnArgGlyLysValAlaSerPheAlaGlnGly 750
Db 2680 GTGATCCCTGTG-----AAAGTAAGGGAATGTGGCTAAGGCTATGTGTTTGA 2730
Qy 751 GlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProLysSerVal 770
Db 2731 GAATTTGTTTGTGTAGACAAAGAGCAT-----CGTGTGAGAAGTCCCATTTGTGCTG 2781
Qy 771 ThrLeu 772
Db 2782 AAGCTC 2787

RESULT 11

US-10-238-4334

; Sequence 4334, Application US/10260238

; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.

; APPLICANT: Moughamer, Todd G.

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Provart, Nicholas

; APPLICANT: Rieke, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION

; FILE REFERENCE: 60111-NP

; CURRENT APPLICATION NUMBER: US/10/260,238

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 4334
; LENGTH: 1934
; TYPE: DNA
; ORGANISM: *Triticum aestivum*
US-10-260-238-4334

Alignment Scores:

Pred. No.: 3,61e-97 Length: 1934
Score: 945.00 Matches: 244
Percent Similarity: 53.79% Conservative: 111
Best Local Similarity: 36.97% Mismatches: 221
Query Match: 23.52% Indels: 84
DB: 12 Gaps: 25

US-09-806-767-2 (1-775) x US-10-260-238-4334 (1-1934)

Qy 141 IleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspThrGly 160
Db 3 ATCATCGAGTTGTTGACACCGGATCTGGCCGAGTCAAGAGCTTCAGCGATGTAGG 62
Qy 161 MetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyLysSerPheSerSer 180
Db 63 TAGGACCGGTCCACCGAGGTGGAAAGGGAGTGCACACCGCGCGGACTGGGGCATC 122
Qy 181 SerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200
Db 123 AACAACTGCAGCGGAAAAATCATCGCGCGGCTTCTAC-----ACCGCT 167
Qy 201 AsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThr 220
Db 168 GGGGTCCCGACGAG-----TACCTCAAAAGGACTCGCTGTCACCGCGGACCACTAC 221
Qy 221 GlyHisGlyThrHisThrAlaSerThrValGlyLysSerSerValSerMet----- 237
Db 222 AGCCACGGCACCATTTGCTTCCACCGAGCTGCTCAGCGTAGGCTGAGCGGCGGAG 281
Qy 238 ---AlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHis 256
Db 282 GAGGCTAGTTCCACGGCTTCTCGGGGGTGGCGGGAGGCGCACCGCGCGCTCGC 341
Qy 257 IleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeuAla 276
Db 342 ATCGGGGTACAGTCCATCTTTGCTGCTACCGTACCGGTATTCAGCACTGTGTCGCGCC 401
Qy 277 IleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPhePro 296
Db 402 ATTGATGACCAATCCACGACGGGTGGATGTATTGCTGCTGCTCCTC-----ACT 452
Qy 297 IleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIle 316
Db 453 GTGCGG---TAGAGAAC-----TCGTACGGTGGCTGCGACCGGTCCAGAGGGGATC 503
Qy 317 SerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAla 336
Db 504 ACTGTGTATCTCCGGGGAACACGAGGACCCAGCGCGCGACACCGTCGAGGATACCTCG 563
Qy 337 ProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValValArg 356
Db 564 CTTGGGTCTCTCAGTACCGCGGAGCAAGATCGACCGCTCTTCCCGACGGTGTATCAG 623
Qy 357 LeuAlaAsnGlyLysLeuLeuTyrGlyLysLeuTyrProGlyLysGlyIleLysAsn 376
Db 624 TTGGGAACAACAGCAAAATGTTGGCCAGTCTCTCTAT-----TACCAAGTGAAGAAC 677
Qy 377 AlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPheCys 396
Db 678 ATCAACTCATCATCAC-----ACCGTAGAAATTTACAAAGCTTATATGT 725
Qy 397 LeuArgGlySerLeuProArgGlu-----IleArgGlyLysMet 410

```

Db      726 GCACAACTCTGTGCGGACACATGAGTTCCTTCATCATATTCAGCTGAAGGGGTATC 785
Qy      411 ValileCys-----AspArgGlyValAsnGlyArgSerGluLysGlyGluAlaVal 427
Db      786 CTGGTGTCTTAAATGATATGATGACGGGATCAATCCATCATGATGATTTCCACCATGCGGG 845
Qy      428 Lys-----GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln 444
Db      846 AAGTACGTGTGGAAGCGGGGATCGGGCTCATCTTCGCG-----CAATACACAACG 899
Qy      445 GluGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyThrGluSer 464
Db      900 GACACCTGAGTCCACAGTGGCTTACTGCCGGG-----ATGCGCTGC 944
Qy      465 ValLeuLeuLysAlaThrValAsnAlaThrVal-----Lys 476
Db      945 GTTATCGTCGACATCTACATCTCCAGGAGATATGCCATACGATTCGACTCAAGCTCG 1004
Qy      477 ProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSer---ArgAlaProGlu 495
Db      1005 CCCACACAAAATCGAACCCACCGCCGTCATGAGCAAGATGTGTGCTTCGAA 1064
Qy      496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
Db      1065 GTGGCGTCTCTCTTCGAGAGGCCATCGCTGATTACCGGCCCTTCATCAGCCTGAC 1124
Qy      516 MetIleAlaProGlyValAsnIleIleAlaAlaThrProGlnAsnLeuGlyProThrGly 535
Db      1125 GTAGCTGGCGCGGAGCCACATCTTAGCAGCA-----GTGGGAGAT----- 1166
Qy      536 LeuProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCys 555
Db      1167 -----TCTTACGTAACTAAGTCAGGAACGCTCCATGGCAGCC 1202
Qy      556 ProHisValSerGlyIleThrAlaLeuLeuArgSerAlaThrProAsnTrpSerProAla 575
Db      1203 CCGCATCGACGGGTGTGCGCGCGCTGCTCAAGCTCTGCACCCGGATGGTCTCTGCT 1262
Qy      576 AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIle 595
Db      1263 GCATAAAGTCAGCCATCGTCACCATCGCATGTACTGATGACGACGCGATCGCGATC 1322
Qy      596 ---LysAspGly-----AsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisVal 612
Db      1323 CTAGCGGAAGGACTACCGCGGAGGTGCGCGATCCATTCGACTATGAGCGCGGAACATC 1382
Qy      613 AsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIle 632
Db      1383 AACCCATCGGGGCGAGCTCATCCGGGCTAATTTATGACATTCATCCACACGATTAACA 1442
Qy      633 ThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsn 652
Db      1443 AAATTCCTT-----CAATCGGCAATGTCAAGAGGGGATATCTCGGATC----- 1487
Qy      653 ValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIle 672
Db      1488 GCAGCTGCAAC-----ACACCTCTATGCCAGAGTATAACCTCAATCGCGTCCATC 1541
Qy      673 AlaValIlePheLysArgGlyLysThrGluMetIleThrArgArgValThrAsnVal 692
Db      1542 TCGGTTCGGATTTTAGAGCGCGATCACC-----GTCTCGAGGACGGTAACCAATGTC 1595
Qy      693 GlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLysValIle 712
Db      1596 GCGAGCTCCATCTGTGTACACCGTGTAGTCCAGACCGCGACTGGAGTCAATATGGAG 1655
Qy      713 ValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTrp 732
Db      1656 GTTTTCCCGAGTGTCTGGTGTTCGATGCTCCGACAAAGTCCGGAATTCGAGGTGAAG 1715
Qy      733 PheValLeuLysLysLysAsnArgGlyLysValAlaSerPheAlaGlnGlyGlnLeu 752

```

```

Db      1716 CTGTCACTCTCTGGAACCTTTCAAGGGGGC-----TACAGTTTGCAGCATC 1763
Qy      753 ThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerValThrLeu 772
Db      1764 ACTTGGCAGATGACCGA-----CAAGTAGTAGGATTCAGTTGGACCCGATTT 1814
RESULT 12
US-09-754-853A-4/c
; Sequence 4, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 4
; LENGTH: 513509
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (111805)..(113968), (114684)..(115204)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(513509)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 318013_region_A3
US-09-754-853A-4
Alignment Scores:
Pred. No.: 5,25e-89 Length: 513509
Score: 909.00 Matches: 285
Percent Similarity: 29.98% Conservative: 108
Best Local Similarity: 21.74% Mismatches: 216
Query Match: 22.62% Indels: 702
DB: 11 Gaps: 22
US-09-806-767-2 (1-775) x US-09-754-853A-4 (1-513509)
Qy      145 LeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetProSerIle 164
Db      195773 TTAACGCTAGGTGTGTGGCCCTGAATCAAGAGCTTTAGTGAGCAAGGTTTGGGACCAATT 195714
Qy      165 ProArgLysTrpLysGlyIleCysGlnGluGly----- 175
Db      195713 CCATCAAGTGGAGAGGAAATCTGTGATAACGGAATTCATATCTTTTCACCTGCAACAGG 195654
Qy      175 ----- 175
Db      195653 TACCATCTTCCTTAGAATGGAATTAACACTATTAACTATTTGTGAGTCATTTATAATGGA 195594
Qy      175 ----- 175
Db      195593 TATAATAAAAATGTTTCTTATCTAATTTGTTTAACTTCTGTAGTCATATGCTTTAAATTTG 195534
Qy      175 ----- 175
Db      195533 GCACCTCTTTATAGGTGAATACACTAATGAAGGGTTAAGCACCTCTTTGAACGACAGACA 195474
Qy      175 ----- 175
Db      195473 TGTTCCTTTTCGTAFTCTTATCTTTATAGTACTAGTGTTCACACCAACCAACCAACCAAT 195414
Qy      176 -----GlusPheSerSerSerSerCysAsnArg 185

```


Qy 493 -----AlaProGluValAlaGlnPheSerAlaArgProSerLeuAlaAsnProSer 510
Db 193208 AAGCCAGCCCACTTATGGAGGATTTTCATCAAAAGGACCAATACTATGTCGCAGAG 193149
Qy 511 IleLeuLys
Db 193148 ATCCATAAAGGTGATTAACCTACAGTGTATTGGCTTGTAAATTCATGTAATGTGTAG 193089
Qy 514 -----ProAspVetIleAlaProGlyVal 521
Db 193088 ATATCTAAGAACTCTTAATTTGTCCAAATATATTCAGCCTGATATCATCTGCACAGCGGTG 193029
Qy 522 AsnIleAlaAlaLysProGlnAsnLeuGlyProThrGlyLeuProTyTrAspSerArg 541
Db 193028 TCAGTTATAGAGCCTTATCTAGAGCCCAAGGACCAACCAATCAAGTGTTCGACAGCCG 192969
Qy 542 ArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIle 561
Db 192968 AGGATTCGGTTTAACTCAGTCTCAGGCACATCAATGTCATGCCCTCATGTTTCAGGCATT 192909
Qy 562 ThrAlaLeuIleArgSerAlaTyProAsnTrpSerProAlaAlaIleLysSerAlaLeu 581
Db 192908 GTGGAGCTGTAGAGCCTTATATCTCTACGTGAGTACCGCTGCAATTAATCAGCAATC 192849
Qy 582 MetThrAlaAsp
Db 192848 ATGACCACAGGTGAGTGAGTGTTCTTCATGGTTTCTCAGCTTCAAACTCTTTT 192789
Qy 586 -----
Db 192788 TCAGCCAACTACAATAAATGTCACGAAATATAAAGTTAGATATATAGAAAATA 192729
Qy 586 -----
Db 192728 ACTTAGATATAGTTTCTTAATTTCTGCTAGTGTATAAAGGTTTTCATTAAAGATTGT 192669
Qy 586 -----
Db 192668 GCAAGTGAAACTCTAATTTCTAATTTTAAAAAAGGATTTACAGTACCTAAGAAACT 192609
Qy 587 -----LeuTyAsnArgGlnGlyValAlaIleLysAspGly 598
Db 192608 ATATGTAAGTATGTAAGTTATGTTCTTAAAGATTTATGAGGGAGGTGAAACTCCAATTCT 192549
Qy 599 Asn
Db 192548 AATTGGAGACGACCACTAGATATCTAAGGAATGTATCTAAGTTTGTCCATACATAT 192489
Qy 599 -----
Db 192488 ATAGAAAGCAATTTGATTAGTGAGTAAAGTAGCTATCAAAAATTTGATGTTTGTCT 192429
Qy 600 -----LysPro-----601
Db 192428 CTGCTACTTTCAGTACACACTAGACAAATGAGTGAGCCACTACTGAAATGCTACTGAT 192369
Qy 602 -----AlaGlyValPheAlaIleGlyValAlaGlnHisValAsnProGlnLysAlaIleAsn 619
Db 192368 GCGAAGGCGACACCATTCAGTTACGGGCGAGGATGTTCAACCAACAGAGCAATGGAT 192309
Qy 620 ProGlyLeuValTyAsnIleGlnProValAspTyIleThrTyLeuCysThrLeuGly 639
Db 192308 CCCGGCTAGTTTATGACATAACAAATGATGATTACCTCAACTCTTATGCTGTTAGGA 192249
Qy 640 PheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeu 659
Db 192248 TACAACGACGACACAGATCTCAGTGTTCACAGAGTCCATATAAATGCCG-----192198
Qy 660 ArgLysAsnProGlyPheSerLeuAsnTyProSerIleAlaValIlePheLysArgGly 679
Db 192197 AAGAAATTTAGTCTCTCAACCTGAACATATCCCTCAATCACTACAGTCCCAAACTCTCTGGG 192138
Qy 680 LysThrThruMetIleThrArgValThrAsnValGlySerProAsnSerIleTy 699

Db 192137 TCAGTGACA-----GTTACAGGACATTTGAAAAAATGTTGTTCTCCAGGACT---TAC 192087
Qy 700 SerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuVal 719
Db 192086 ATTGCTCATGTTCAAAACCCGCTATGGAATCACCGTTTCTGTGAAGCCACGATCTTGAAG 192027
Qy 720 PheLysHisValAspGlnThrLeuSerTyArgValTrpPheValLeuLysLysAsn 739
Db 192026 TTCAAAATGTTGTGAAGAGAGAGAGCTTTAAATTAACCTTC-----NAGGCCATG 191976
Qy 740 ArgGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsn 759
Db 191975 CAAGAAAGGCCCAATAACTACGCGTTTGGAAACTGATATGTCAGATGGCAAGCAT 191916
Qy 760 LeuMetGlnArgValArgSerProIleSerVal 770
Db 191915 TAC-----GTCAAGAGTCCAATAGTTGTG 191892
RESULT 13
US-09-938-842A-64
; Sequence 64, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kresps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: SPRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 64
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-64
Alignment Scores:
Pred. No.: 5, 8e-92 Length: 2451
Score: 901.00 Matches: 269
Percent Similarity: 48.25% Conservative: 130
Best Local Similarity: 32.53% Mismatches: 327
Query Match: 22.42% Indels: 101
DB: 10 Gaps: 28
US-09-806-767-2 (1-775) x US-09-938-842A-64 (1-2451)
Qy 6 PhePheLeuCysIle-----IlePheLeuLeuPheCysSerSer 19
Db 28 TTTTTCATGTTCTTCTCCTCACTGTGCACAGACAGATTACATGTCATGTAAGAGGAGAA 87
Qy 20 SerGluIleLeuGlnLys-----GlnThrTyIleValGlnLeuHis 33
Db 88 CCAATCATTAGTTACAAAGTGGTGATAATGGAATTTGAAGCAACTGCTGTGAATCTCAT 147
Qy 34 ProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeuSerPheLeu 53
Db 148 GAGAAAAATGATGATACAACTGACATCAATGTCATGTCATGTCATGTCATGTCATGTCAT 207
Qy 54 GlnGluAlaValLeuGlyValGluGluGluGluGluGluGluGluGluGluGluGluGlu 73
Db 208 CATGATATGCTTCTTGAATG-----CTCTTTGTGGAGGATCTTACAAAAGCTTTAC 261
Qy 74 SerTyGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluIle 93

Db 262 AGCTATAAACACCTCATTAATGTTTGCAGCTCATGTATCCCTCATGACGGCGGAAATG 321
Qy 94 LeuArgTyrSerProGluValAlaValArgProAspHisValLeuGlnValGlnThr 113
Db 322 CTTCGTGCGGCGCTGTGTGAATCTGTGTAGATAGAGATTGGAAGATTAGGAACCTACT 381
Qy 114 ThrTyrSerTyrIlePheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrp----- 131
Db 382 ACACATACACACAGTTTGGGATTACCA-----ACTGATGTTTGGCCCACT 429
Qy 132 -----SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal 149
Db 430 GGTGCTGTTATGATAGACGAGGGAAGATATTGTTATGCTTTATTGACTCAGGGATA 489
Qy 150 TrpProGluSerProSerPhe-----AspAspThrClyMetProSerIleProArg--- 166
Db 490 TTTCACATACACCAAGTTTTCCTCTCACACACACACAGTATGATGGCCCTCATCT 549
Qy 167 LysTrpLysGlyIleCysGlnGlyGluSerPheSerSerSerCysAsnArgLys 186
Db 550 AGTTACAAAGGAATGTGAAGAGATCCTCATACTAAGATCAGTTTTCATGATGGAG 609
Qy 187 LeuIleGlyAlaArgPheIleArgGlyHisArgValAlaAsnSerProGluGluSer 206
Db 610 ATCATAGGCGCAGCATTTGCTGAAGCCGTAAAGCAGCTGGTGT-----TTCAAT 663
Qy 207 ProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThr 226
Db 664 CCGGATATC-----GACTTGTCTTCCCAATGGATGGATGGATGGATGGATGGATGGAT 717
Qy 227 AlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly 246
Db 718 GCAGCTATTGCGAGTGGGAACAATGGTATTCAGTGGAGATGCGAGTATGAGTTTGG 777
Qy 247 ValAlaArgGlyMetAlaProGlyValAlaHisIleAlaValTyrLysValCysTrp----- 284
Db 778 AAGCAAGCGGATGGCTCTCTCGTGAAGGATGCTGTTTACAAGCTCTTACCGACT 837
Qy 265 PheAsnGlyCysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLys 284
Db 838 TTTGGAGC---TTGTAGCTGATGTGTGTGCTGCCATTGATCAGGCTGTTATGATGGA 894
Qy 285 ValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIle 304
Db 895 GTAGATATTGTAGCTCTCTCGTGTGTCCTCAACAGCTCTCCAGCTACTACAAAGACAACA 954
Qy 305 AlaIleGlyThrPhe-----ArgAlaMetGluArgGlyIleSerValIle 319
Db 955 TTCTTGATCCATTGTGATGCTACTCTTGGGCTGTAAAGCTGGTGTGTTTGTGCT 1014
Qy 320 CysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpVal 339
Db 1015 CAAGCTGCTGGAACGCGGTTCTCTTCCGAAACTCTGTTTCATACACCCCTTGGATA 1074
Qy 340 SerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValAlaArgLeuAlaAsn 359
Db 1075 ACTACTGTGCTGTGCAATGTATGATCGCAGATACAAAAATCATCTGACTCTTGGAAAT 1134
Qy 360 GlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArg 379
Db 1135 GGTAAATGTCTGCGGAATAGGATATTCTCTTACTCGACCTCATCTGTTTCATACAAG 1194
Qy 380 GluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPhe----- 395
Db 1195 ATGTTTCTGCAATGATGTTCTGCTGTTCTGTTCTGTTATGAAATACAATCCGTCGGAC 1254
Qy 396 CysLeuArgGlySer---LeuProArgGluGluIleArgGlyLysMetValIleCysAsp 414
Db 1255 TGCCAGAGCCAGAGTCTTGAACAAAAATTTAGTCGAGGAAACATCTGCTTTTGGGA 1314
Qy 415 ArgGlyValAsn-----GlyArgSerGluLysGlyGluAlaValLys 428

Db 1315 TATTCCTTTCAATTTGCTGTTGCTGCTTCCAGTTCCTCATCAAGAAAAGTTGCTGAACCTGCCAG 1374
Qy 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleLeuAsnGlnGluGluAspSe 448
Db 1375 CATCTAGCGCTGCTGTTGCTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCT 1434
Qy 448 IleAspValHisIleLeu---LeuProAlaThrLeuIle---GlyTyrThrGluSerValLeu 466
Db 1435 GATCCGCTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1494
Qy 467 LeuLysAlaTyrValAsnAlaThrVal----- 475
Db 1495 TTGATGATTACTACAATGTCAACAGTCAAGAGATTGGATGGGAAGGTTAAAGGACTTT 1554
Qy 476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db 1555 AAAGCTGAAGAGCATCGGAGAGGTTTGAACCCATTCTTCAAAATCCGACCTGAA 1614
Qy 496 ValAlaGlnPheSerAlaArgGlyPro-----SerLeuAlaAsnProSer 510
Db 1615 GTTCTCTGTTCTCAGCTCGAGGACCCCAATACCAAGATTTTCAAGTTCAGGATGCTGAT 1674
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaIleAlaIleProGlnAsn 530
Db 1675 CTTCCTCAAAACAGATATTTCTGCTCCAGGCTCTTAAATATGCTGCTGCTGCTGCTGCTGCTGCT 1734
Qy 531 LeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMetSerGly 550
Db 1735 ---GGAACAGACGAGGCTAATATATCGCGAA-----GGATTGCTCAATTTCTGCG 1785
Qy 551 ThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaIlePro 570
Db 1786 ACACAGATGGCTGCACACACATTCGCGGTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1845
Qy 571 AsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArg 590
Db 1846 CAATGGAGTCCAGCTGCCATTAATCAGCTTTGATGACGCTTCAACAGCTCATAGATAGA 1905
Qy 591 GlnGlyLysAlaIleLysAspGlyAsn-----LysPro 601
Db 1906 GCAGGAAGACCTCTCCAAGCACCAATATTCTGAACAGACAGACAGTAACTTTGTTAAA 1965
Qy 602 AlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGly 621
Db 1966 GCAACTCGTGTGATATGGAAGGCTCATGCTCAATCCAGCGCTGCTTCAACCTGCT 2025
Qy 622 LeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeu---GlyPhe 640
Db 2026 CTCATCTTGATGCGAGGTTATGAGGATATATAGGTTCTTGTGCACACACCTGTTATC 2085
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 2086 GATGCTCAGAGATAAAAAAATTCA-----AACACTCCATGCAATTTCAAAATGTT 2139
Qy 661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 2140 CATCTCTCAAACTTC-----AACACGCTCCATAGCCATC-----TCTCATCTCGTC 2187
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSer 700
Db 2188 AGAACACAAACCGTAACAGAGAGTACGATGTTGCGGAAGAGAGAGAACATACACA 2247
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db 2248 ATCATCATCGAGATGGAGCCCAATCCCATCGAAGTACCTCTCTCTGCAATGACAGTA 2307
Qy 721 Lys---HisValAspGlnThrLeuSer----- 728
Db 2308 AGACGCGCGCTTCTAGAACCTTTTCCGTGACTTAACAGTACAGTACATCAGTACGAGCT 2367
Qy 729 TyrArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAla-----Ser 746
Db 2368 TATAGCTCGGACAGTTACATTGAAGGAAGCGGAGGCAATAAAGTACCTCTCCCTGTG 2427

Qy 747 PheAlaGlnGlyGln 751
 Db 2428 GTTGCTATGGACAA 2442

RESULT 14
 US-09-938-842A-64
 ; Sequence 64, Application US/09938842A
 ; Publication No. US20040009476A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kieps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SRIPI300-3
 ; CURRENT APPLICATION NUMBER: US/09/938,842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264,647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300,111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 64
 ; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-938-842A-64

Alignment Scores:
 Pred. No.: 5,8e-92 Length: 2451
 Score: 201.00 Matches: 269
 Percent Similarity: 48.25% Conservative: 130
 Best Local Similarity: 32.53% Mismatches: 327
 Query Match: 22.42% Indels: 101
 DB: 12 Gaps: 28

US-09-806-767-2 (1-775) x US-09-938-842A-64 (1-2451)

Qy 6 PhePheLeuCysIle-----IlePheLeuLeuPheCysSerSer 19
 Db 28 TTTTTCATGTTTCTCCTCACTGTGACAGCAGAGATTTTACATTTGCTGACTATGGAGAGAA 87
 Qy 20 SerGluLeuGlnLys-----GlnThrTyrIleValGlnLeuHis 33
 Db 88 CCAATCATTAGTTTCAAAAGGTGTGATAATGGATTGGAACAACTGCTGTGGAACTGAT 147
 Qy 34 ProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeuSerPheLeu 53
 Db 148 GAGAAATTTGATATACAGTGAACCTGGTGTGATCATCATACCCCGTCCACCTTGAGAGAA 207
 Qy 54 GlnGluAlaValLeuGlyValGluGluGluGluGluProSerArgLeuLeuTyr 73
 Db 208 CATGATATGCTTCTTGAATG-----CTCTTTGTGAAGGATCTTACAAAAGCTTTAC 261
 Qy 74 SerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluIle 93
 Db 262 AGCTATAAACACCTCATTAAATGTTTTCAGCTCATGTATCCCTCATCAGCGGGAATG 321
 Qy 94 LeuArgTyrSerProGluValValAlaValArgProAspHisValLeuGlnValGlnThr 113
 Db 322 CTTCGTCGCGGCTGTGTGAAATCTGTGATAGAGATTGGAAGTTAGGAACCTTACT 381
 Qy 114 ThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTyr----- 131
 Db 382 ACACATACACACAGTTTGGGATTACCA-----ACTGATGTTTGGCAACT 429
 Qy 132 -----SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal 149
 Db 430 GGTGGTGGTTATGATAGACAGCGGGAAGATATTGTTTGGCTTTATTGACTCAGGGATA 489

Qy 150 TrpProGluSerProSerPhe-----AspAspThrGlyMetProSerIleProArg--- 166
 Db 490 TTTCCACATCACCACAGTTTGGCTCTCACACACACAGTACCTTATGCGCCCTCATCCT 549
 Qy 167 LysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLys 186
 Db 550 AGTTACAAAGGGAATGTGAAGAGATCCTCATACATAGATCAGTTTCTGCAATGGGAAG 609
 Qy 187 LeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSer 206
 Db 610 ATCATAGAGCGCAGCATTTTGTGAAGCGCTAAAGCAGCTGTGTCT-----TTCAAT 663
 Qy 207 ProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThr 226
 Db 664 CCGGATATC-----GACTTTGCTTCCCAATGATGCGATGACATGGAAGTACACACA 717
 Qy 227 AlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly 246
 Db 718 GCAGCTATTGCAGCTGGGAACAATGGTATTCAGTGGAGATGCACGGTTATGAGTTTGGGA 777
 Qy 247 ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrp----- 264
 Db 778 AAAGCAAGCGGATGGCTCTCGTCAAGGATTCGTGTTTACAAAGCTCTCTACCGACTT 837
 Qy 265 PheAsnGlyCysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLys 284
 Db 838 TTTGAGGC---TTTGTAGCTGATGTGTGCTGCCATTGATCAGGCTGTTCATGATGGA 894
 Qy 285 ValAspValLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIle 304
 Db 895 GTAGATATTTTGAAGTCTCTGGTTGGTCCAAACAGTCTCTCAGCTACTACAAAGACAACA 954
 Qy 305 AlaIleGlyThrPhe-----ArgAlaMetGluArgGlyIleSerValIle 319
 Db 955 TTCTGTAATCATTGATGCTACACTCTGGGCTGTAAAAGCTGTGTTTGTGCT 1014
 Qy 320 CysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpVal 339
 Db 1015 CAAGCTGCTGGAACGGAGTCCCTTTCCGAAACTCTGTTTTCATACAGCCCTTGGATA 1074
 Qy 340 SerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsn 359
 Db 1075 ACTACTGTGGTGTGCAATGATGATCGCAGATACAAAATCATCTGACTCTTGGAAAT 1134
 Qy 360 GlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArg 379
 Db 1135 GGTAAATGCTTGGCGGAATAGGATATCTCTTCTACTCGACCTCATCTCATACAG 1194
 Qy 380 GluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPhe----- 395
 Db 1195 ATGTTTCTGCAATGATGTTCTGTTGTTCTTGTGTAATAAATCAATCCGTCGGAC 1254
 Qy 396 CysLeuArgGlySer---LeuProArgGluGluIleArgGlyLysMetValIleCysAsp 414
 Db 1255 TGCCAGAAAGCCAGAGTCTTGAACAAAATTAAGTCGAAGGAAACATCTCTGTTTGGGA 1314
 Qy 415 ArgGlyValAsn-----GlyArgSerGluLysGlyGluAlaValLys 428
 Db 1315 TATCTTTTCAATTTGCTGTGCTGCTGCTCCATCAAGAAAGTTGTGGAACCTGCCAAG 1374
 Qy 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsn-GlnGluGluAspSe 448
 Db 1375 CATCTAGCGGCTGCTGTTGTTCTGTTGTTGTCGAAACAGTTTCTCCAGGACAAATTC 1434
 Qy 448 rIleAspValHisLeu---LeuProAlaThrLeuIle---GlyTyrThrGluSerValLeu 466
 Db 1435 GATCCCGCTCTCTTGTGATTCCTGGGATTCGATTACAGATGTCTCTAAGTCAATGAT 1494
 Qy 467 LeuLysAlaTyrValAsnAlaThrVal----- 475
 Db 1495 TTGATTGATTACTACAATGTCAACAGTCAAGAGATTGGATGGGAAGGTAAGGACTTT 1554

Qy 476 LysProLysAlaArgIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db 1355 AAAGCTGAAGAGACGCGAGACGGTTTGGAAACCCATTCTTCAAAATCCGACCTGAA 1614
Qy 496 ValAlaGlnPheSerAlaArgGlyPro-----SerLeuAlaAsnProSer 510
Db 1615 GTTGCTCTGTCTCAGCTCGAGGACCAATACCAAGATTTCAGCTTCAGGATGCTGAT 1674
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleAlaAlaTppProGlnAsn 530
Db 1675 CTTCTCAACACGATATCTTGCTCCAGGCTCTTAATATGCTGCTGCTGCTGCAAT 1734
Qy 531 LeuGlyProThrGlyLeuProTyroAspSerArgValAsnPheThrValMetSerGly 550
Db 1735 ---GGAACAGACGAGGCTAACTATATCGCGAA-----GGAATTGCACTAAATTTCTGGC 1785
Qy 551 ThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyroPro 570
Db 1786 ACAAGCATGGCTGCACACACATATCGGGTATTCTGCTGTGTAACAGACGATCCT 1845
Qy 571 AsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyroAspArg 590
Db 1846 CAATGGAGTCCAGCTGCATTAATCACTTTGTATGACGACTTCAACAGTCATAGATAGA 1905
Qy 591 GlnGlyLysAlaIleLysAspGlyAsn-----LysPro 601
Db 1906 GCAGGAGACCTCTCCAGACACACAATATCTGAACAGACAGACAGTAACACTTGTATA 1965
Qy 602 AlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGly 621
Db 1966 GCAACTCGGTTTGATATATGAAGCGGTCTGTCAATCAAGCGGTCTGTAGACCTGGT 2025
Qy 622 LeuValTyroAsnIleGlnProValAspTyroIleThrTyroCysThrLeu---GlyPhe 640
Db 2026 CTCATCTTGATGAGGTTATGAGGATATATAGGTTCTTGTGCACACACACCTGGTATC 2085
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 2086 GATGCTCAGGAGTAAACAACTTCACA-----AACACTCCATGCAATTTCAAAATGGTT 2139
Qy 661 LysAsnProGlyPheSerLeuAsnTyroProSerIleAlaValIlePheLysArgGlyLys 680
Db 2140 CATCTTCAAACTTC-----AACACGCCATCCATAGCCATC-----TCTCATCTGTC 2187
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyroSer 700
Db 2188 AGAACACAAACCGTAACACAGAGAGTGCAGAAATGTCGGAAGAGAGAAAGAACATACACA 2247
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db 2248 ATCATCTGAGGATGAGCCAGCAATCGCATCGAAGTGAGTCTCTCTGCAATGACAGTA 2307
Qy 721 Lys---HisValAspGlnThrLeuSer----- 728
Db 2308 AGACGGCGGCTTCTAGACCTTTTCGGTGACTCTACAGTGCAGTGCAGTGGAGCT 2367
Qy 729 TyrArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAla-----Ser 746
Db 2368 TATAGCTTCGACAGGTTACATTGAAGGAGCGGCGGATTAAGTGACTCTCTCTGCTG 2427
Qy 747 PheAlaGlnGlyGln 751
Db 2428 GTTGCTATGGGACAA 2442

RESULT 15

US-10-260-238-901
; Sequence 901, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Krebs, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 901
; LENGTH: 2018
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (442)..(442)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-901

Alignment Scores:

Pred. No.:	9,5e-80	Length:	2018
Score:	793.00	Matches:	248
Percent Similarity:	47.42%	Conservative:	129
Best Local Similarity:	31.19%	Mismatches:	254
Query Match:	19.74%	Indels:	165
DB:	12	Gaps:	27

US-09-806-767-2 (1-775) x US-10-260-238-901 (1-2018)

Qy	10	IleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLysGlnThrTyrlle	29
Db	6	ATTATATTCTTCTTCTCGTGCATTTCTACTACTTACTTACTTTTGCAC-----ATTGTT	56
Qy	30	ValGln-----LeuHisProAsnSerGluThrAlaLysThr---	41
Db	57	GTTTCAGACACACGGATCTCGTAAGCTTTTACATAGCATATCTCGGTGATGTAACACAG	116
Qy	42	-----PheAlaSerLysPheAspTrpHisLeuSerPheLeuGln	54
Db	117	CCACCCCGACGATGTCATCGCTTCACACCATGATATGCTCACCACGTCTCTTGG-----	170
Qy	55	GluAlaValLeuGlyValGluGluGluGluProSerSerArgLeuLeuTyroSer	74
Db	171	-----AAGAGCAAGGAAGATTCTTTGGGCTCTTATCATCCACAAC	209
Qy	75	TyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGlu-IleLe	94
Db	210	TACAAGCATGTTCTCAGGCTTCGCCGTGATGTTACTGAAGATCAAGCAAGAACTT	269
Qy	94	uArgTyroSerProGluValValAlaValArgProAspHisValLeuGlnValGlnThrTh	114
Db	270	GCAGGAATTCGGGAAGTCTTAGTGTGAACCAAGCAAAACGTATATACGCGGCGACAC	329
Qy	114	rTyroSerTyroLysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSe	134
Db	330	TCGGAGCTGGGACATGCTTGGACTGAACACTACAGATGCCACTGAACGTCTCCAGCGAAC	389
Qy	134	rArgPheGlyGlnGlyThrIleGlyValLeuAspThr-GlyValTrpProGluSerP	154
Db	390	CAACTACGCGGAGGAGATAATCATTTGGCATTTGTTACACTGGGATC-----	436
Qy	154	roSerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnG	174
Db	437	-----TGCCANG	443

QY 174 luGlyGluSerPheSerSerCysAsnArgLysLeuIleGlyAlaAaGpPheI 194
Db 444 TCGCGCAGAGATGGCGCAGCAAAAT---GAGCGCAGATATCCGCGCGGTTA--- 496
QY 194 leArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArgGluTyrI 214
Db 497 -----CACGCCGGGTGGACGAGGACGACCTCAAGATCGACTACC 536
QY 214 leSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerS 234
Db 537 TCTCGCCCGCGCAGCGCAGCGCATGGCACGCGCATCGCGCGCGCGCGGCTCGG 596
QY 234 erValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProG 254
Db 597 TCGGGAG-----GCGGGGTCCGGAGC----- 619
QY 254 lyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIle 274
Db 620 -----ACGCCACCGTGC 632
QY 274 euAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeuGlyG 294
Db 633 TGGCGCCATCGACGACGCGATCCACGACGCGGTGGACGCTGCTCCTCCCTGGAGA 692
QY 294 lyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluA 314
Db 693 AC-----TCCTTCGCGCGCGACGACGCGCGTCCAGA 722
QY 314 rgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaA 334
Db 723 AGGA-----TACCGTGTTACGCGCCATGAACCTTGGCCTGGCCTCAGGTGGTTCAGA 779
QY 334 snThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaV 354
Db 780 ACATCGCCCTTGGGTATCATCCGTCGCGGCGACGACATCGATCGCTCCTCCGACGG 839
QY 354 alValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyr---ProGlyLysG 373
Db 840 TGATCATGTTGGGTGACAGCAGACAGATAGTGGACAACTCTCTACTCTCAAGGGAAGA 899
QY 373 lyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlys 393
Db 900 ACTCATCGCTCAGCGGTTTCAGACGCTTGTCTGTGGAGTT-----GGAG 944
QY 393 erGluPheCysLeuArgGlySerLeuProArgGlu-GluIleArgGlyLysMetValIle 412
Db 945 CTGAATGG-----CACGGATGTCAAAGGACGATTTGCTG 980
QY 413 Cys-----AspArgGlyValAsnGlyArgSerGluLysGlyGluAla 426
Db 981 TGGCATCTTTTACTCTCAACAAACCATCTATCTTTTCCAAAGAAGCTCTAGGAACGTC 1040
QY 427 ValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGlu 446
Db 1041 GTGAAA---GGCGGAGGGGTGGCATGATTTTCGTTCAGTATACGTGGGATATTGAAGT 1097
QY 447 AspSerIleAspValHisLeuLeuProAlaThrLeuIleGly---TyrThrGluSerVal 465
Db 1098 TCGACAGCCAGATGCAACGGTATTGCGTGCTTATTGCGACTATTACAGATGAACACAG 1157
QY 466 LeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGly 485
Db 1158 ATCGGAATAATACATCTACTGTCGAGAGTTCCACCATTTGAGATCGATCCA-GCAGCG 1216
QY 486 ThrValIleGlyArgSer---ArgAlaProGluValAlaGlnPheSerAlaArgGlyPro 504
Db 1217 ACCGTAAAGGGGAATGAGATAATGGCACCAAAAGTGGCAGATTCTCTTCAAGAGGCCCT 1276
QY 505 SerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIle 524
Db 1277 TCAACTGATTACCTGAAATATTCAAGCCAGACATAGTGCACCGAGATTCAACATCTTA 1336

QY 525 AlaAlaTrpProGlnAsnLeuGlyProThrThrGlyLeuProTyrAspSerArgValAsn 544
Db 1337 GCAGCA-----GTGAAGGCACA 1354
QY 545 PheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeu 564
Db 1355 TACGCATTGTGCTCAGGACGTCAATGGCTACTCCGCATGTAGCAGCGTTGTGCGCGCTG 1414
QY 565 IleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThr 584
Db 1415 CTGAAGCTTGCATCCCAAGCTGTACCTGCTGCATTAATCAGCATAGTCACCCACA 1474
QY 585 -AlaAspLeuTyrAspArgGlnGlyLysAlaIle---LysAspGly-----AsnLysPr 601
Db 1475 GGCATCTGTACTGATGAGCGTGCGCATGCCCATACTGCGCAGAGGGTTGCTCGGAAGAT 1534
QY 601 oAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGl 621
Db 1535 TGCTGACCATTTGACTATGAGCGGCGGCATCAACCCCTAACAGACGAGCTGATCTCTGG 1594
QY 621 yLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeu---CysThrLeuGlyPh 640
Db 1595 CCTGATTTCAGACATTGATCCACGCGATTACAAACAAATCTTTGGATGCACCTGTA----- 1649
QY 640 eThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuAr 660
Db 1650 -----AAGCCATATGTAAGTGCAATGCG-----AC 1675
QY 660 GlyAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLy 680
Db 1676 ATCGTTGCTGGGTATTATCTGAACCTACCGTCCATCTCAGTCCCGATCTTAGGTATCC 1735
QY 680 sThrThrGluMetIleThrArgValThrAsnValGlySerProAsnSerIleTyrSe 700
Db 1736 GGTC-----GTCGTATCAAGACGCGTCACAAATGTAGCTGAGTTGATGCTGTTTACCA 1789
QY 700 rValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPh 720
Db 1790 CGCTGCAATCGAGAGCCCACTGCGGTAAGATGGATGTTGAGCCACCTGTTCTGGTGTT 1849
QY 720 eLysHisValAspGlnThrLeuSerTyrArgValTrpPheValLeuLysLysLysAsnAr 740
Db 1850 CAATGCTGCAACAAAGTTTCATACGTTCCAGTTAAGCTTTCACCTCTATGGAAGTGCA 1909
QY 740 gGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLe 760
Db 1910 GGGG-----GACTACAGCTTGGGAGCCTTACTTGG-----CACAAACGG 1948
QY 760 uMetGlnhrgValArgSerProIleSerValThrLeu 772
Db 1949 CCAGAAAGACAGTGAGAAATCCCGATAGCAGTCAGGATT 1985

Search completed: February 3, 2004, 06:34:23
Job time : 1710 secs